



République Algérienne Démocratique et Populaire  
Ministère de l'enseignement Supérieur et de la Recherche Scientifique

**UNIVERSITE SAAD DAHLEB BLIDA 1**

Faculté des Sciences de la Nature et de la Vie  
Département de Biologie Physiologie Cellulaire  
Mémoire de fin d'études en vue de l'obtention du diplôme  
de Master en science de la nature et de la vie  
Filière Sciences Biologiques

**Option : Génétique**

**Thème :**

**Etude de validation d'un protocole de séquençage des régions hypervariables de l'ADN mitochondrial humain référence dans le cadre de l'identification génétique**

**Réalisé par :** M<sup>lle</sup> MENARI Sana  
M<sup>lle</sup> KHALFI Ahlem Houria

**Devant le jury composé de :**

Nom Prénom	Grade	Lieu	Qualité
M <sup>r</sup> BENYAHIA N.	Maitre-assistant A	Université de Blida 1	-Président
M <sup>me</sup> AISSANI R.	Maitre de conférences B	Université de Blida 1	-Examinatrice
M <sup>me</sup> KEBBAS S.	Maitre de conférences A	Université de Blida 1	-Promotrice
M <sup>lle</sup> BESSEKRI M.W	MSC science forensique	INCC-GN	-Co-promotrice
M <sup>me</sup> TAHRI K.			-Invitée
M <sup>r</sup> MOHAMED SAID R.			-Invité

Promotion 2018-2019

## ***Dédicaces***

*Avec l'expression de ma reconnaissance je dédie ce modeste travail :*

*A l'homme de ma vie, mon exemple éternel, mon soutien moral et ma source de joie et de bonheur, celui qui s'est sacrifié pour me voir heureuse, à toi mon cher père, tu as toujours été à mes côtés pour m'encourager et me soutenir que dieu t'accorde bonne santé et longue vie*

*A la femme qui m'a toujours couvert avec sa tendresse, qui n'a jamais dit non à mes exigences et qui n'a épargné aucun effort pour me rendre heureuse, mon amour éternel, ma source de force et de courage , à toi mon adorable mère que dieu te garde pour moi*

*Quoi que je fasse ou que je dise, je ne saurai point exprimer mon respect et ma gratitude comme il se doit, j'espère que vos bénédictions m'accompagnent toujours, je vous aime énormément.*

*A ma binôme et ma chère amie Sana, la source de mon sourire, qui m'a accompagné durant mon cursus universitaire et pendant l'élaboration de ce travail et qui était toujours là pour moi, merci beaucoup pour tout.*

*A mon aimable amie Sarah qui m'a aidé, qui m'a encouragé et qui m'a accompagné durant mes meilleurs et pires moments, mille merci à toi.*

*Ainsi qu'à tous les membres de ma famille pour leur aide , leur soutien et leurs encouragements durant la réalisation de ce travail.*

*Ahlem Houria*

## *Dédicaces*

*Je dédie ce travail :*

*A ma chère mère qui m'a appris à ne jamais laisser tomber, je la remercie pour son soutien, ses sacrifices, ses encouragements tout au long de mon cursus et surtout pour son amour sans limite, qui m'ont amené à ce que je suis aujourd'hui.*

*A ma sœur Manel et mon frère Mehdi qui ont fait que cette année soit remplie de joie et de rire.*

*A mon amie et binôme Ahlem, je la remercie pour son soutien moral, son amitié, sa présence et pour tout ce qu'elle m'a apporté depuis notre rencontre.*

*A mes grands-parents maternels et ma tante Fatiha qui m'ont épaulé tout au long de mes études avec leur présence et leur amour.*

*A mes amis Nahla, Sami, Leila, Lilya, Doha, Hiba, Wassim et Ramzi je les remercie pour leurs encouragements et leur présence à mes côtés.*

*Que ce travail soit un enrichissement et source de savoir.*

*Sana*

## Remerciements

Nous exprimons notre pleine gratitude à notre promotrice **Mme KEBBAS S.** pour avoir accepté de diriger ce travail, d'avoir répondu à nos questions, pour les connaissances qu'elle nous a rapportées, sa supervision éclairée et sa présence tout au long de la rédaction du mémoire ainsi que pour ses remarques constructives et ses encouragements.

Nous tenons à exprimer toutes nos reconnaissances à l'égard de **Mme BESSEKRI M.W.** pour nous avoir encadré, orienté et guidé durant toute la période de notre travail ainsi que pour sa patience, sa disponibilité, ses judicieux conseils qui ont contribué à compléter nos connaissances et le partage de ses compétences.

Nos remerciements les plus sincères s'adressent aux membres du jury **Mr BENYAHIA N.** et **Mme AISSANI R.** pour l'intérêt qu'ils ont porté à notre travail et d'avoir accepté de l'examiner et de l'enrichir avec leurs connaissances.

Nos vifs remerciements s'adressent à **Mme TAHRI K.** pour son aide, sa gentillesse et le temps qu'elle nous a consacré.

Nous désirons remercier **Mr MOHAMED SAID R.** pour son aide, son soutien et pour son intervention qui nous a permis d'avoir l'opportunité de réaliser ce travail.

Nos remerciements distingués s'adressent à **Mr BOUSSOUFI A.** Chef de Département de Biologie de L'INCC-GN et **Mr LAZREG S.** Responsable du laboratoire pour la confiance qu'ils nous ont accordée et pour nous avoir permis d'accomplir ce mémoire dans les meilleures conditions possibles.

Nous remercions également tout le personnel du laboratoire pour leur gentillesse, leur modestie ainsi que leur disponibilité.

Enfin, nous tenons à remercier toutes les personnes qui ont contribué de près ou de loin pour l'élaboration de ce modeste travail.

## Résumé

En criminalistique l'état de l'échantillon de l'ADN est un point clé pour aboutir à une identification fiable. Pour pallier au soucis rencontrés avec l'ADN nucléaire en analyses, l'utilisation de l'ADNmt dans ce domaine est devenue incontournable.

Dans le but d'élaborer un protocole pour l'identification génétique par l'ADNmt, notre travail porte sur le séquençage des régions hypervariables de la région de contrôle (D-Loop) d'un ADN mitochondrial de source connue (non dégradé) à partir d'échantillons biologiques de type salive.

Notre étude s'est déroulée au sein du laboratoire du département de biologie de l'INCC de la Gendarmerie Nationale d'Alger. Toutes les étapes du protocole ont été analysées tour à tour dont les principales étant l'extraction de l'ADN mitochondrial, son amplification, son séquençage en suivant trois différents designs de primers. La lecture et l'interprétation des résultats ont également fait l'objet de l'analyse en utilisant deux logiciels le Sequencher® v5.4.2 et le Sequencing Analysis Software v6.0 Applied Biosystems™.

A travers l'élaboration de ce protocole nous avons pu, d'une part obtenir le mitotype de chaque échantillon analysé et d'autre part contribuer à sa validation en utilisant différents critères et différentes procédures. Nous avons eu aussi la possibilité d'appuyer notre protocole et les différents choix en se référant à des études précédentes. Enfin notre étude nous a permis d'améliorer le protocole afin d'avoir de meilleurs résultats d'analyses.

**Mots clés :** Identification génétique, ADNmt, HVI, HVII, HVIII, D-Loop, Séquençage, Amorces, Mitotypes.

## Abstract

In forensic science the state of the DNA sample is a key point to achieve reliable identification. To overcome the problems encountered with nuclear DNA in analyzes, the use of mtDNA in this field has become unavoidable.

In order to develop a protocol for genetic identification by mtDNA, our work focuses on the sequencing of hypervariable regions of the control region (D-Loop) of mtDNA from known source (non-degraded) from saliva-type biological samples. .

Our study took place in the laboratory of the Biology Department of the INCC of the National Gendarmerie of Algiers. All the steps of the protocol were analyzed successively, the main ones being the extraction of the mitochondrial DNA, its amplification, its sequencing by following three different primers designs. The reading and interpretation of the results were also analyzed using two software packages, Sequencher® v5.4.2 and Sequencing Analysis Software v6.0 Applied Biosystems™.

Through the development of this protocol we were able, on the one hand, to obtain the mitotype of each sample analyzed and on the other hand to contribute to its validation using different criteria and different procedures. We also had the opportunity to confirm our protocol and the different choices by referring to previous studies. Finally, our study allowed us to improve the protocol in order to have better analysis results.

**Key words:** Genetic identification, mtDNA, HVI, HVII, HVIII, D-Loop, Sequencing, Primers, Mitotypes.

## المخلص

في علم الطب الشرعي ، تعتبر حالة عينة الحمض النووي نقطة أساسية لتحديد هوية موثوق به. للتغلب على المشاكل التي واجهها الحمض النووي النووي في التحاليل ، أصبح استخدام الحمض النووي الميتوكوندري في هذا المجال أمرًا لا مفر منه.

لغرض تطوير بروتوكول لتحديد الجينات بواسطة الحمض النووي الميتوكوندري ، يركز عملنا على تسلسل المناطق شديدة التغير في منطقة التحكم (D-Loop) من الحمض النووي الميتوكوندري من مصدر معروف (غير المتحلل) من العينات البيولوجية من نوع اللعاب .

تم إجراء دراستنا في مختبر قسم الأحياء التابع للمجلس للمعهد الوطني للأدلة الجنائية و علم الإجرام للدرك الوطني بالجزائر العاصمة. تم تحليل جميع خطوات البروتوكول بدورها ، أهمها استخراج الحمض النووي الميتوكوندري ، تضخيمه ، تسلسله باتباع ثلاثة تصميمات أساسية مختلفة. كما تم تحليل النتائج وتفسيرها باستخدام نوعين من البرامج ، Sequencing Analysis Software v6.0 Applied Biosystems™ و Sequencher® v5.4.2 .

من خلال تطوير هذا البروتوكول ، تمكنا ، من ناحية ، من الحصول على نموذج مصغر لكل عينة تم تحليلها ومن ناحية أخرى للمساهمة في التحقق من صحتها باستخدام معايير مختلفة وإجراءات مختلفة. كما أتاحت لنا الفرصة لدعم بروتوكولنا والخيارات المختلفة من خلال الرجوع إلى الدراسات السابقة. وأخيرا ، سمحت لنا دراستنا لتحسين البروتوكول من أجل الحصول على نتائج تحليل أفضل.

**الكلمات المفتاحية:** التعرف الوراثي ، الحمض النووي الميتوكوندري ، HVI ، HVII ، HVIII ، D-Loop ، التسلسل ، البادئات ، أنماط الميتوتيب.

## Liste des Abréviations

**ADN** : Acide Désoxyribonucléique

**ADNmt** : Acide Désoxyribonucléique Mitochondrial

**ADNnuc** : Acide Désoxyribonucléique Nucléaire

**ADN pol** : ADN polymérase

**CR** : Région de Contrôle

**CRS** : Cambridge Reference Sequence (séquence de référence de cambridge)

**ddNTP** : Didésoxyribonucléotide Triphosphate

**D-Loop** : Displacement Loop (boucle de déplacement)

**dNTPs** : Didésoxyribonucléotide

**EDNAP** : European DNA Profiling Group (groupe européen du profilage d'ADN)

**EMPOP** : European DNA Profiling group (EDNAP) mtDNA population Database ( la base de donnée de l'ADN mitochondrial du groupe EDNAP)

**F** : Primer Forward (sens)

**HLA** : Human Leukocyte Antigen

**HSP** : Heavy Strand Promoter (promoteur du brin lourd)

**HVI** : Région Hypervariable I

**HVII** : Région Hypervariable II

**HVIII** : Région Hypervariable III

**LSP** : Light Strand Promoter (promoteur du brin léger)

**LINE** : Long Interspersed Nuclear Element

**LTR** : Long Terminal Repeats

**NIST** : National Institute of Standards and Technology

**O<sub>H</sub>** : Origine de réplication du brin lourd

**O<sub>L</sub>** : Origine de réplication du brin léger



**pb** : Paire de Base

**PCR**: Polymerase Chain Reaction (réaction de polymérisation en chaîne)

**R** : Primer Reverse (anti-sens)

**rCRS** : revised Cambridge Reference Sequence (séquence de référence de cambridge révisée)

**ROS** : Reactive Oxygen Species

**RFU**: unites de fluorescence relatives

**SINE** : Short Interspersed Nuclear Element

**SNP** : Single Nucleotide Polymorphism (polymorphisme mono-nucléotidique)

**STR**: Short Tandem Repeat (courte répétition en tandem)

**SWGAM** : Scientific Working Group on DNA Analysis Methods

**VNTR** : Variable Number Tandem Repeats

## Liste des figures

<b>Figure 1.</b> Les 13 loci STR de la base de données CODIS (Kobilinsky et al., 2005) .....	4
<b>Figure 2.</b> Génome de la mitochondrie (Nicholls et Minczuk, 2014) .....	7
<b>Figure 3.</b> Les haplogroupes Majeurs de la population Humaine (Stewart et Chinnery, 2015).....	11
<b>Figure 4.</b> Diagramme représentatif des étapes du protocole d'analyse de l'ADNmt .....	15
<b>Figure 5.</b> Les étapes de la PCR (NCBI, 2017).....	19
<b>Figure 6.</b> Thermocycleur Gene Amp PCR System 9700 - Applied biosystems (Marshall scientific) .	20
<b>Figure 7.</b> Exemple d'un plan de dépôt sur gel utilisé .....	21
<b>Figure 8.</b> La réaction de séquençage de Sanger (Claire Peltier, Futura-Sciences, 2011).....	23
<b>Figure 9.</b> Primers de la 1 <sup>ère</sup> stratégie et leurs positions .....	24
<b>Figure 10.</b> Primers de la 2 <sup>ème</sup> stratégie et leurs positions.....	24
<b>Figure 11.</b> Primers de la 3 <sup>ème</sup> stratégie et leurs positions.....	25
<b>Figure 12.</b> Primers de la 4 <sup>ème</sup> stratégie et leurs positions.....	26
<b>Figure 13.</b> Exemple d'un plan de plaque utilisé .....	28
<b>Figure 14.</b> Placement de la plaque sur l'agitateur (Applied Biosystems BigDye® XTerminator™ Purification Kit, Protocol, 2007).....	29
<b>Figure 15.</b> Séquenceur 3500xl Genetic Analyzer d'Applied Biosystems. ....	30
<b>Figure 16.</b> Exemple d'électrophorèse sur gel d'agarose.....	33
<b>Figure 17.</b> Exemples de résultats d'électrophorégramme obtenus après une lecture des séquences analysées.....	36
<b>Figure 18.</b> Exemple de la lecture des résultats sous forme de « Raw Data » .....	37
<b>Figure 19.</b> Exemple de résultat de séquençage de HVI du brin sens et HVII du brin antisens de l'échantillon E631 contenant un « Poly C » court dans les deux régions. ....	38
<b>Figure 20.</b> Exemple de résultat de séquençage de HVI du brin sens et HVII du brin antisens de l'échantillon E630 contenant un « Poly C » court et un allongé .....	39
<b>Figure 21.</b> Exemple de résultat de séquençage de HVI du brin sens et HVII du brin antisens de l'échantillon E231 contenant un « Poly C » allongé dans les régions.....	40
<b>Figure 22.</b> Raw Data de l'échantillon E215 dilué et non dilué séquençé avec l'amorce R16410.....	41
<b>Figure 23.</b> Electrophorégramme du T- F15971.....	42
<b>Figure 24.</b> Données brutes « Raw Data » du T-2 F15971 .....	42
<b>Figure 25.</b> La CR du T <sup>+</sup> assemblée.....	44
<b>Figure 26.</b> La CR du E630-A assemblée .....	45
<b>Figure 27.</b> La CR du E231-A. ....	46
<b>Figure 28.</b> Un alignement d'une séquence d'un échantillon avec la « rCRS ». ....	47
<b>Figure 29.</b> Hétéroplasmie du E115-A. ....	48
<b>Figure 30.</b> Hétéroplasmie du E195. ....	48
<b>Figure 31.</b> Profil du E215 validant le critère de répétabilité.....	52
<b>Figure 32.</b> Profil du E631 validant le critère de reproductibilité.....	53
<b>Figure 33.</b> Séquence de la région de contrôle (D-loop).....	33
<b>Figure 34.</b> Différences entre la « rCRS » et l'ADNmt des deux échantillons du « SRM 2392 » au niveau des régions hypervariables (non-codantes) (Levin et al., 2003) .....	36
<b>Figure 35.</b> Profil du T <sup>+</sup> (9947A).....	37
<b>Figure 36.</b> Profil du C <sup>+</sup> (CHR).....	38

## Liste des tableaux

<b>Tableau I.</b> Différences nucléotidiques entre la CRS et la rCRS .....	9
<b>Tableau II.</b> Les amorces d'amplification utilisés.....	18
<b>Tableau III.</b> Calculs des volumes du mélange réactionnel de l'amplification .....	18
<b>Tableau IV.</b> Programme thermocycleur pour l'amplification.....	19
<b>Tableau V.</b> Calculs des volumes du mélange réactionnel de la réaction de séquençage .....	26
<b>Tableau VI.</b> Programme du Thermocycleur pour le séquençage.....	27
<b>Tableau VII.</b> Calculs des volumes du mélange réactionnel de la solution de purification.....	28
<b>Tableau VIII.</b> Paramètres du Séquenceur 3500xl spécifique au Protocole étudié.....	30
<b>Tableau IX.</b> Résultats de la quantification des produits amplifiés.....	34
<b>Tableau X.</b> Résultats de la quantification des échantillons sans lavage et avec lavage.....	35
<b>Tableau XI.</b> Résultats de séquençage lu par « SeqA ».....	43
<b>Tableau XII.</b> Résultat du profil du T <sup>+</sup> (9947A). .....	49
<b>Tableau XIII.</b> Résultat du profil du C <sup>+</sup> (CHR). .....	50
<b>Tableau XIV.</b> Moyennes des valeurs de l'intensité du signal (RFU) de chaque base des témoins négatifs.....	51
<b>Tableau XV.</b> Récapitulatif des profils des échantillons étudiés.....	54
<b>Tableau XVI.</b> Matériels non Biologique.....	34
<b>Tableau XVII.</b> Les amorces utilisées en séquençage.....	35

# Sommaire

Introduction .....	1
Chapitre I : Etude bibliographique .....	2
I.1. L'identification génétique .....	2
I.2. Polymorphismes génétiques humains .....	2
I.2.1. Nature des régions variables de l'ADN.....	3
I.2.3. Les STRs autosomaux .....	3
I.2.4. Les STRs du chromosome Y .....	4
I.2.5. Les SNPs .....	5
I.2.6. Les régions hypervariables de l'ADN mitochondrial .....	5
I.3. L'ADN mitochondrial en identification génétique .....	6
I.3.1. Structure et composition de l'ADN mitochondrial .....	6
I.3.2. Caractéristiques de l'ADN mitochondrial et son rôle en identification génétique .....	8
I.4. Caractéristiques des haplotypes de l'ADN mitochondrial .....	9
I.4.1. La séquence d'Anderson .....	9
I.4.2. Hétéroplasmie de l'ADNmt .....	10
I.4.3. Les haplogroupes et la base de données EMPOP .....	10
I.5. Validation des méthodes d'analyse.....	12
I.5.1. Définition de la validation.....	12
I.5.2. Objectifs de la validation.....	12
I.5.3. Critères de la validation.....	13
Chapitre II : Matériels et méthodes .....	14
II.1. Recommandation pour une analyse de l'ADNmt Référence.....	16
II.2 Echantillonnage .....	16
II.3. Extraction.....	16
II.3.1. Principe de la FTA .....	16
II.3.2. Poinçonnage et lavage.....	17
II.4. Amplification par PCR .....	17
II.4.1. Principe .....	17
II.4.2. Choix des amorces .....	17
II.4.3. Amplification par le kit « AmpliTaq Gold 360 Master Mix ».....	18
II.5. contrôle de qualité par électrophorèse sur gel .....	20
II.5.1. Principe .....	20
II.5.2. Préparation du gel pour 180ml.....	20

II.5.3. Dépôt et migration sur gel.....	21
II.5.4. Visualisation des bandes et estimation de la quantité.....	21
II.6. Purification post-amplification .....	21
II.6.1. Principe .....	21
II.6.2. Purification à l'ExoSAP-IT®.....	21
II.7. Dilution.....	22
II.8. Réaction de Séquençage .....	22
II.8.1. Principe de la méthode.....	22
II.8.2. Mode opératoire .....	23
II.9. Purification Post-Séquençage .....	27
II.9.1. Principe .....	27
II.9.2. Mode opératoire .....	27
II.10. Electrophorèse Capillaire.....	29
II.10.1. Principe .....	29
II.10.2. Mode opératoire .....	29
II.11. Lecture des Données.....	31
II.11.1.1. Principe.....	31
II.11.1.2. Mode opératoire .....	31
II.11.2.1. Principe.....	31
II.11.2.2 Mode opératoire .....	31
II.12. Paramètre de validation .....	32
Chapitre III : Résultats et discussion.....	33
III.1. Résultats de la quantification .....	33
III.2. Résultats du séquençage.....	35
III.2.1. Sequencing Analysis Software v6.0 « SeqA » .....	35
III.2.2. Résultats du Sequencher .....	43
III.3. Résultats de l'étude des témoins négatifs.....	51
III.4. Résultats de la validation.....	52
III.5. Discussion Générale .....	56
Références bibliographiques .....	60
Annexes	

# Introduction

## **Introduction**

La criminalistique est un domaine impliquant l'interprétation scientifique des preuves en appliquant des connaissances et des méthodes scientifiques pour la résolution des affaires juridiques. L'identification génétique est un point crucial en criminalistique qui demande des techniques précises, fiables et pouvant donner des résultats souhaités. Ceci dans le but d'établir des profils d'ADN en vue d'identifier des individus et/ou de déterminer leur filiation ([Min Shen et Duarte Nuno Vieira, 2016](#)).

En effet, chaque individu possède un génome unique et propre à lui, présent dans chacune des cellules de son organisme. Les séquences variables du génome autrement dit les marqueurs génétiques, constituent ce qu'on appelle un polymorphisme. Les techniques les plus utilisées pour l'identification génétique reposent sur l'analyse du polymorphisme de courtes séquences de nucléotides répétées en tandem appelées STR (Short Tandem Repeats). Au quotidien chaque individu laisse des traces biologiques contenant des particules d'ADN. En sciences forensiques, ces traces sont recueillies et les analyses des marqueurs permettent d'obtenir un profil ADN type rendant ainsi possible l'identification génétique ([Pham-Hoai et al., 2017](#)).

Connaissant la nature sensible de l'ADN humain, il est souvent difficile de l'extraire et de le manipuler ce qui ne garantit pas l'obtention de profils fiables et discriminant. Cependant différentes techniques d'analyse d'ADN ont été améliorées et affinées notamment l'analyse de l'ADN mitochondrial. Cette technique est avantageuse, elle repose sur les différentes caractéristiques de l'ADNmt. Sa présence en plusieurs copies, son taux élevé de mutations ainsi que sa transmission à travers la lignée maternelle sans occurrence de la recombinaison, ont fait de lui un outil de choix pour l'identification génétique, en particulier celle de la lignée maternelle ([Wallace et Chalkia, 2013](#)).

Dans notre présente étude, nous avons contribué à l'établissement et la validation d'un protocole qui s'appuie sur le séquençage des régions hypervariables (HVI, HVII et HVIII) de l'ADNmt de provenance connue (non dégradé). Ceci en cernant des paramètres spécifiques tel que la concentration et la pureté de l'ADNmt et en affinant le processus d'analyse afin de les adapter en fonction de la présence ou l'absence de « poly C allongé » de l'échantillon, dans le but de l'identification génétique.

# **Chapitre I : Etude bibliographique**



## Chapitre I : Etude bibliographique

### I.1. L'identification génétique

Les sciences forensiques ou criminalistiques sont un ensemble de disciplines qui servent le système juridique par des méthodes d'analyses scientifiques pointues permettant d'identifier la nature et la source des différents indices et traces retrouvés sur les scènes de crimes, et par conséquent l'identification des individus impliqués. La génétique forensique compte parmi les spécialités les plus précises et les plus requises par la justice, notamment pour l'élucidation des affaires de meurtres, de viols et de découvertes de cadavres (Edlund, 2010).

Au début du 20<sup>e</sup> siècle, la distinction biologique était fondée sur l'utilisation des marqueurs génétiques sérologiques classiques, qui se faisait par l'étude des groupes sanguins, la détermination des antigènes HLA (système HLA) et l'analyse du polymorphisme des enzymes et des protéines sériques (Carracedo, 2005 ; Edlund, 2010). L'identification des individus a évolué et est devenue plus fine et précise, avec la découverte de l'empreinte génétique par Alec Jeffreys en 1985 (Jeffreys, 2005). Dès lors, l'utilité de traiter ces informations génétiques trouvées dans l'ADN en tant que marqueurs médico-légaux est devenue évidente.

Suite à sa découverte, Jeffreys a introduit la méthode de « DNA finger printing » qui se base sur l'analyse des régions portant de très grandes variabilités de l'ADN, présentes à certains endroits non codants du génome humain (Jeffreys, 2005). Le même auteur a démontré aussi que plus le nombre de régions analysées dans le chromosome est élevé, moins est la probabilité que deux individus aient la même variabilité d'ADN au même endroit, introduisant ainsi la notion du profil génétique (Cole et Prainsack, 2015).

Cette nouvelle technique, ainsi que l'affinement de l'idée de base mise au point, ont permis de comparer le patrimoine génétique des personnes et de déterminer si deux profils d'ADN proviendrait de la même personne. Ceci a permis aux enquêteurs sur les criminels d'établir des liens entre des personnes suspectes et des traces biologiques laissées sur les scènes de crimes (Prainsack et Aronson, 2015).

### I.2. Polymorphismes génétiques humains

Depuis la découverte de sa structure en 1953 par Rosalind Franklin, Watson et Crick, l'ADN n'a cessé d'être étudié et a souvent été décrit comme étant a souvent été décrit par le terme « blueprint » désignant une sorte de plan de la vie car il contient toute l'information nécessaire à un organisme pour fonctionner et se reproduire ( Goodwin et al, 2007).

Chez l'être humain l'information génétique est portée dans un génome d'environ 3 milliards de paires de base. Moins de 10% de ce génome correspond aux régions codantes appelées « gènes », qui contiennent les informations requises à la synthèse des protéines par une cellule ; alors que plus de 90% représente les régions non codantes appelées autrefois « Junk DNA » (Panneerchelvam et Norazmi, 2003). Dans environ 30% de l'ADN non codant, il existe de nombreux allèles différents portant des variations génétiques de séquences et de longueurs, ces dernières sont sous forme d'unités répétitives et variables en nombres de paires de bases. Cette diversité génétique est la conséquence de l'accumulation des mutations au cours de l'évolution

dans les régions non codantes de l'ADN car celles-ci ne sont pas contrôlées par la pression de sélection, contrairement aux régions codantes, où ce phénomène joue un rôle dans le maintien des fonctions spécifiques des gènes (Panneerchelvam et Norazmi, 2003; Butler, 2005; Carracedo, 2005; Butler *et al.*, 2007).

### **I.2.1. Nature des régions variables de l'ADN**

Le polymorphisme génétique est représenté par les régions variables de l'ADN, celles-ci sont largement composées de séquences répétées, dont deux classes sont bien distinguées. La 1ère classe étant les séquences répétées dispersées, elles représentent 45% du génome, dérivées d'éléments mobiles ou transposables et sont constituées des transposons et des rétrotransposons (LTRs, SINEs et LINEs) (Goodwin *et al.*, 2007). La seconde classe a été mise en évidence par l'étude de Sir Alec Jeffreys en 1984, cette classe est dite « Séquence Répétée en Tandem », qui sont une succession de motifs d'ADN de longueurs différentes allant de quelques nucléotides à plusieurs centaines, répétées les unes derrière les autres et dispersées dans le génome à des endroits particuliers. Ces séquences sont considérées comme des marqueurs génétiques pouvant constituer un profil ADN unique d'un individu (Goodwin *et al.*, 2007; Carracedo, 2005).

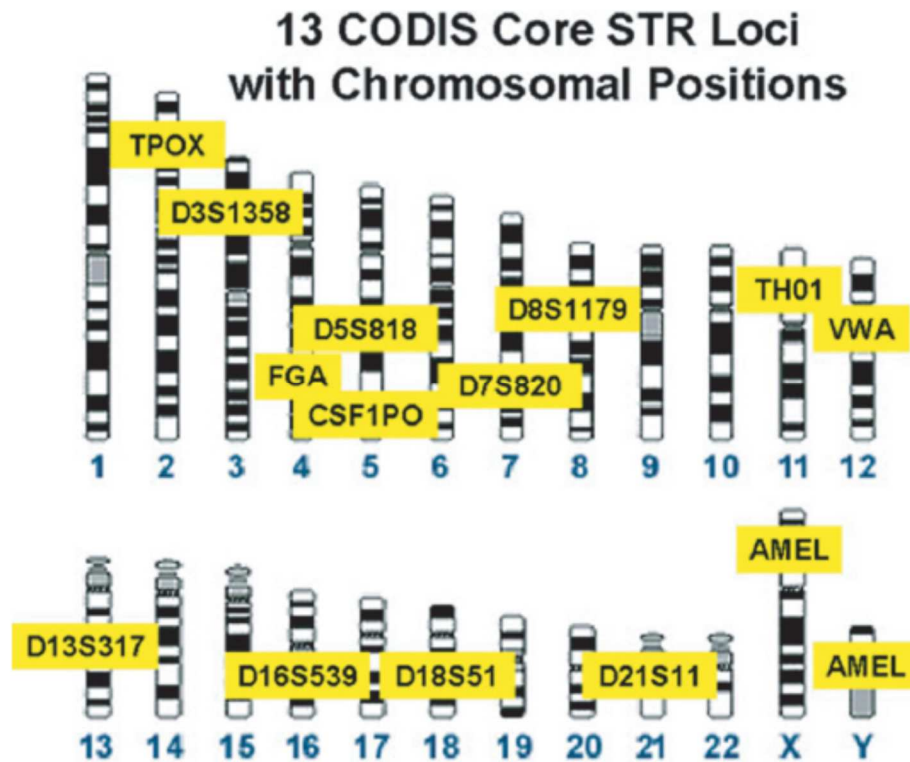
Initialement les méthodes d'identification génétique se basaient sur les minisatellites ou VNTR (Variable Number Tandem Repeats), dont les unités de répétition comptent entre 20 à 100 pb répétées jusqu'à cent fois; leur utilisation pour l'analyse était très discriminative, cependant, leur grande taille fut à l'origine de complications lors de l'analyse des échantillons (Runa et Walsh, 2006). De ce fait et en plus de l'avènement du processus de la PCR (Polymerase Chain Reaction) permettant d'amplifier des régions spécifiques de l'ADN, l'utilisation d'autres marqueurs génétiques à caractère polymorphe devient possible. Ces marqueurs sont représentés par les microsatellites, également appelé STR (Short Tandem Repeats) dont les unités de répétition sont plus courtes et donc plus faciles à caractériser, ainsi que les SNPs (Single Nucleotide Polymorphisms) représentant des mutations ponctuelles distinguables (Goodwin *et al.*, 2007; Edlund, 2010).

### **I.2.3. Les STRs autosomaux**

Les STRs (Short Tandem Repeats) ou microsatellites, sont des séquences répétées hautement polymorphiques, constituées d'un motif nucléotidique allant de 2 à 6 paires de bases, qui représente l'unité de base se répétant un certain nombre de fois (polymorphisme de longueur) et sont trouvées majoritairement dans les régions non codantes de l'ADN (Kobilinsky *et al.*, 2005; Dumache *et al.*, 2016). Les STRs présentent différents allèles qui varient par le nombre de répétitions des unités. Un individu peut être homozygote (même nombre de répétitions) ou hétérozygote (différent nombre de répétitions) pour un locus particulier (Edlund, 2010).

Les microsatellites sont les marqueurs génétiques les plus fréquemment utilisés pour établir des profils génétiques, car ils présentent plusieurs caractéristiques avantageuses telles que l'amplification aisée par PCR, leur sensibilité et leur pouvoir discriminant. La majorité des STRs choisis et validés pour l'analyse de routine contiennent des répétitions tétra-nucléotidiques, afin de minimiser les erreurs lors de l'amplification pouvant être dû à des unités répétitives trop courtes ou trop longues (Panneerchelvam et Norazmi, 2003; Goodwin *et al.*, 2007; Dumache *et al.*, 2016).

Les treize loci des STRs de la banque de donnée CODIS sont utilisés communément pour l'identification humaine afin d'unifier les données génétiques, cependant d'autres loci spécifiques supplémentaires sont également utilisés par la communauté européenne ou encore asiatique. Ces marqueurs offrent un fort niveau de discrimination (Panneerchelvam et Norazmi, 2003 ; Edlund, 2010).



**Figure 1.** Les 13 loci STR de la base de données CODIS (Kobilinsky et al., 2005)

La nomenclature des marqueurs de l'ADN se fait selon la position de ces derniers. Si un marqueur fait partie d'un gène, le nom du gène est utilisé pour la désignation dans ce cas : **TH01** : provient du gène humain de la tyrosine hydroxylase et le «01» vient du fait que la région de répétition en question est située dans l'intron **1**. **TPOX** : la peroxydase thyroïdienne, 10ème intron. **CSF1PO** : Proto-oncogène c-fms, 6ème intron. **FGA** : Alpha fibrinogène, 3ème intron. **vWA** : Facteur de von Willebrand, 40ème intron. **AMEL** : Amelogénine, pour déterminer le sexe de l'individu. Les marqueurs situés en dehors du gène sont désignés par leur position chromosomique. « D » signifie ADN. Le caractère suivant fait référence au numéro de chromosome. Le « S » désigne que le marqueur d'ADN est une séquence à copie unique. Le dernier numéro indique l'ordre dans lequel le marqueur a été découvert et classé pour un chromosome particulier. Par exemple le marqueur **D21S11** : **D** : ADN (DNA), **21** : chromosome 21, **S** : copie unique (Single copie), **11** : 11ème décrit sur le chromosome 21.

#### I.2.4. Les STRs du chromosome Y

Le chromosome Y est l'un des deux chromosomes sexuels (l'autre étant le X), présent uniquement chez les mâles. Il est l'un des plus petits chromosomes du génome humain, d'une taille d'environ 60 Mb et ne contient que 161 gènes au niveau du bras court. Sa partie non codante contient plusieurs STRs polymorphes chez les hommes non apparentés et sont hérités à travers la lignée paternelle avec peu de changement au fil des générations, ce qui les rends un

outil supplémentaire de très grande valeur dans l'identification génétique (Kobilinsky et al, 2005 ; Edlund, 2010 ; Butler, 2011).

Environ 400 loci de STRs ont été identifiés sur le chromosome Y. Ces loci sont dit « liés » car ils sont situés sur le même chromosome et les allèles analysés aux différents loci constituent ce qu'on appelle un haplotype (Edlund, 2010 ; Butler, 2011).

Leur intérêt fût majeur en criminalistique pour l'analyse d'ADN dans le cas de comparaison avec des individus de sexes masculins et aussi dans le cas d'analyse de mélanges d'ADN masculins et féminins (cas de viols) ; ainsi que dans d'autres champs d'applications tels que les tests de paternités, l'étude des tendances des migrations des populations humaines à travers l'histoire et la recherche généalogique (Butler, 2011 ; Dumache et al, 2016).

### **I.2.5. Les SNPs**

Considérés comme des marqueurs génétiques, les SNPs (**S**ingle **N**ucleotide **P**olymorphism) ou polymorphismes nucléotidiques, sont des positions ou loci dans le génome, qui présentent une variabilité entre les individus, ne se situant que dans une seule base (polymorphisme de séquence), à un point particulier du génome humain dans les régions codantes et non codantes des autosomes ainsi que des gonosomes. Les SNPs sont très abondants dans le génome et sont généralement bi-alléliques et sont le résultat d'une transition ou d'une transversion (Kobilinsky et al, 2005 ; Runa et Walsh, 2006 ; Butler, 2011).

Du fait de la possibilité de les utiliser en tant que petits amplicons dans le cas d'échantillons dégradés, ainsi que leur faible taux de mutation estimé à  $10^{-8}$ , leur intérêt pour l'identification d'individus fait des SNPs des marqueurs additionnels de choix dans les cas d'identification des restes squelettiques et cadavériques putréfiés (Runa et Walsh, 2006 ; Dumache et al, 2016).

Cependant, leur utilisation dans ce cadre, est limitée par l'incapacité à amplifier simultanément suffisamment de SNPs à partir de petites quantités d'ADN, donc il est nécessaire d'analyser un grand nombre de SNPs afin d'obtenir un fort pouvoir de discrimination (Butler, 2011).

### **I.2.6. Les régions hypervariables de l'ADN mitochondrial**

Le polymorphisme génétique ne concerne pas seulement l'ADN nucléaire mais aussi l'ADN mitochondrial qui se situe dans la matrice mitochondriale.

En effet l'ADNmt est sujet aux mutations plus souvent que l'ADN nucléaire (Butler, 2005 ; Goodwin et al, 2007), au niveau de sa région non codante connue sous le nom de « la région de contrôle ». Cette région contient des portions hautement variables entre les individus en fonction de l'ethnie, probablement en raison des erreurs de répllication de l'ADNmt et des mutations, non corrigées, souvent dues aux produits de la phosphorylation oxydative (**ROS** : reactive oxygen species) (Kobilinsky et al, 2005 ; Goodwin et al, 2007).

En criminalistique, la plupart des études de l'ADNmt pour l'identification ont portées sur les variations trouvées dans deux segments hypervariables de la région de contrôle qui sont nommés « Hypervariable Region I » HVI et « Hypervariable Region II » HVII ; ces deux régions contiennent le plus haut niveau de variation dans le génome mitochondrial. HVI se situe entre les bases 16024 et 16365 (contient 342 pb) alors que HVII se trouve entre les bases 73 et

340 (contient 268 pb) (Kobilinsky *et al.*, 2005 ; Samehsalari et Reddy, 2018). L'étude d'une troisième région « Hypervariable Region III » HVIII se trouvant entre les bases 438 et 576 (contient 138 pb), est devenue indispensable pour obtenir plus d'informations dans le but de l'identification. Ainsi, dans ce type d'investigation certains laboratoires utilisent dans leurs analyses toute la région de contrôle « CRs » (Lutz *et al.*, 2000 ; Butler, 2011).

### I.3. L'ADN mitochondrial en identification génétique

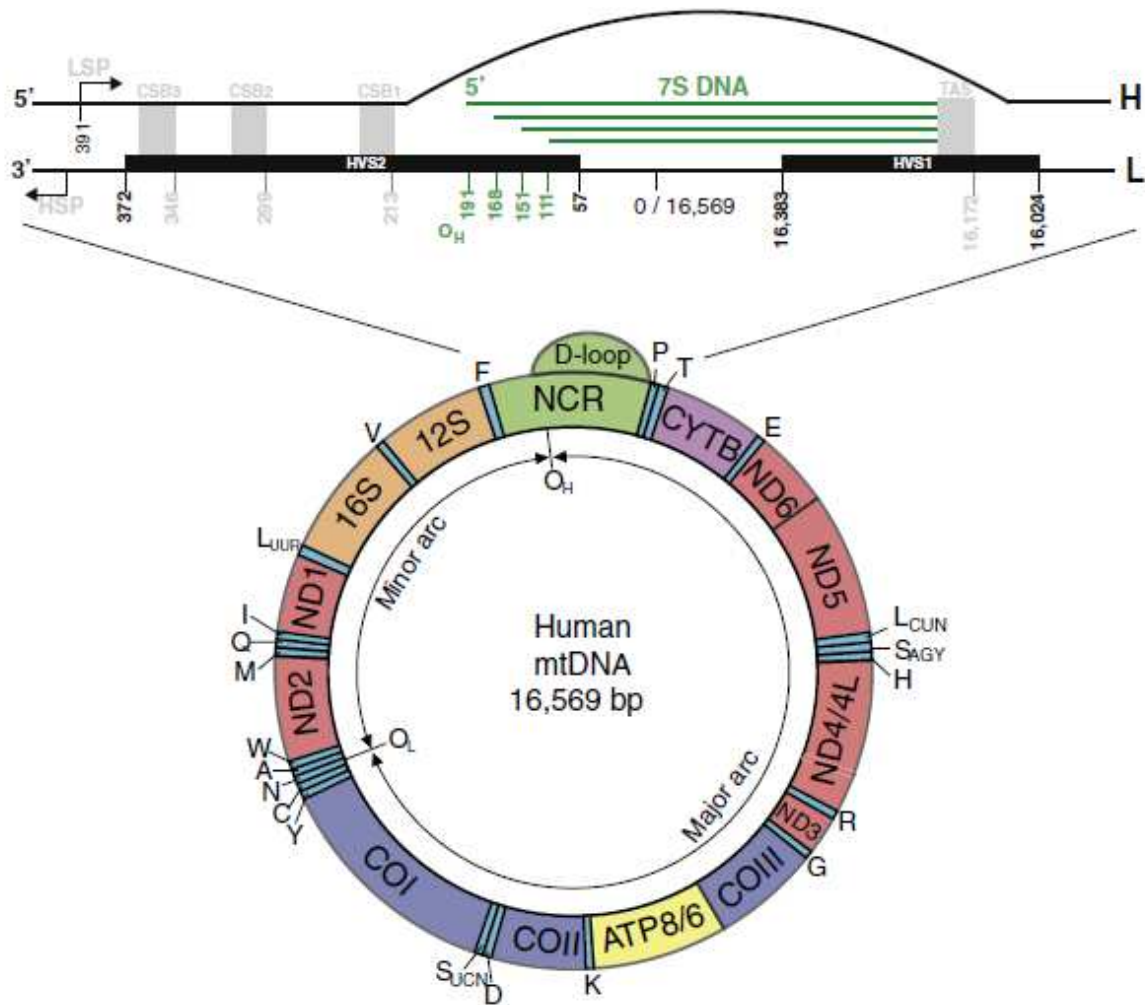
#### I.3.1. Structure et composition de l'ADN mitochondrial

Les mitochondries sont des organites cellulaires se trouvant en grand nombre dans le cytoplasme des cellules eucaryotes (plusieurs centaines par cellule). En premier lieu, elles ont une fonction vitale pour l'organisme, car ces organites produisent l'énergie requise pour la cellule à travers le processus de phosphorylation oxydative lors de la respiration cellulaire. En effet, ce phénomène de la chaîne respiratoire permet le transfert d'électrons à partir de molécules réductrices à des molécules oxydantes. Chez la mitochondrie, la molécule d'O<sub>2</sub> a un fort pouvoir oxydant ce qui la rend comme le meilleur accepteur final d'électron lors de la phosphorylation oxydative. La réduction de l'oxygène (4 électrons) produit de l'H<sub>2</sub>O, ceci ne rend pas la réduction de l'oxygène sans danger. Effectivement si la réduction ne concerne qu'un ou deux électrons ceci permet à l'oxygène de produire respectivement soit de l'ion superoxyde ou de l'ion peroxyde. Ces deux espèces réactives de l'oxygène produisent des radicaux libres potentiellement dangereux car ils oxydent les protéines, les dNTPs (qui, en s'incorporant dans l'ADN créent une instabilité génomique et une diminution de la fidélité de réplication) et l'ADN générant des mutations (Valko *et al.*, 2007 et Kong *et al.*, 2014). En deuxième lieu, elles possèdent un génome indépendant car ces dernières ont abouti de l'endosymbiose entre une cellule eucaryote primitif et une bactérie ancestrale ( $\alpha$ -protéobacterium) il y a de cela 2 milliards d'années approximativement, ce qui donne à ces organites une double membrane (externe et interne), ainsi que la capacité de produire de l'ATP et enfin un génome propre à elle (Taanman, 1999 ; Kobilinsky *et al.*, 2005 ; Goodwin, 2007 ; Friedman et Nunnari, 2014 ; Reid, 2018).

Une mitochondrie contient entre 1 à 15 copies d'ADN circulaire appelé ainsi ADN mitochondrial (ADNmt). La séquence du génome mitochondrial humain a été déterminée pour la première fois en 1981 par Anderson *et al.* L'ADNmt humain est double brin d'une longueur totale de 16569 pb et contient trente-sept gènes. Vingt-deux de ces gènes codent pour l'ARN de transfert (ARNt), deux codent pour l'ARN ribosomal (ARNr) « 12S et 16S » et treize codent pour des protéines impliquées dans le processus de la phosphorylation oxydative. Les deux brins de l'ADNmt connus sous le nom de brin lourd « H » (Heavy) et de brin léger « L » (Light) en fonction de leur composition en bases, étant donné que le brin « H » est riche en purines et code pour vingt-huit produits, alors que le brin « L » est riche en pyrimidines et code pour huit produits. Le génome mitochondrial humain montre une similitude avec le génome bactérien, car il est organisé en complexes nucléoprotéiques appelés nucléoïdes et ses gènes ne contiennent pas d'introns ; aussi les séquences intergéniques sont absentes ou limitées à quelques bases (Anderson *et al.*, 1981 ; Taanman, 1999 ; Holland et Parsons, 1999 ; Butler, 2011 ; Nicholls et Minczuk, 2014).

La région de contrôle (CR) appelée aussi « displacement loop » ou D-loop, représente la région non codante du génome mitochondrial et c'est aussi la région d'intérêt pour l'identification génétique. Elle est constituée approximativement de 1200 pb et renferme les

promoteurs pour la transcription de l'ARN et l'origine de réplication de la chaîne lourde. Le système de numérotation de la séquence de l'ADNmt humain commence près du milieu donc la CR occupe la position entre les bases 16024 et 16569, puis se prolonge de la position 1 à 576 (Holland et Parsons, 1999 ; Butler, 2011).



**Figure 2.** Génome de la mitochondrie (Nicholls et Minczuk, 2014)

La réplication de l'ADNmt humain commence par la synthèse du brin lourd « H » à partir de l'origine de réplication spécifique «  $O_H$  ». Cette synthèse nécessite l'incorporation d'un troisième brin d'ADN d'environ 680 bases appelé l'ADN 7S (7SDNA), ce qui conduit à la formation de la boucle de déplacement ou la D-loop. La synthèse du brin léger (L) débute elle aussi au niveau de son origine de réplication spécifique «  $O_L$  » se situant aux deux tiers (2/3) de la distance génomique de  $O_H$ . La réplication du brin « L » ne commence que lorsque la fourche de réplication de la synthèse du brin « H » passe par «  $O_L$  » ce qui l'active (Anderson *et al*, 1981 ; Nicholls et Minczuk, 2014 ; Yasukawa et Kang, 2018 ; Falkenberg, 2018).

Concernant la transcription de l'ADNmt humain, les deux brins « H » et « L » sont complètement transcrits à partir de leurs promoteurs situés dans la D-loop est sont nommés LSP

pour le promoteur du brin léger « L » et HSP pour le promoteur du brin lourd « H ». Les transcrits primaires subissent une maturation pour donner les ARNr 12S et 16S, les ARNt et un certain nombre d'ARNm présomptifs (Anderson *et al.*, 1981 ; Taanman, 1999).

Le code génétique mitochondrial diffère du code génétique universel en quatre codons. Chez la mitochondrie le codon « UGA » représente l'acide aminé Tryptophane (Trp) au lieu du codons STOP, alors que l'acide aminé Méthionine (Met) est représenté par le codon « AUA » qui code pour l'Isoleucine (Ile) dans le code génétique universel et enfin les codons « AGA » et « AGG » codent pour un codon STOP au lieu de l'Arginine (Arg) (Butler, 2011).

### **I.3.2. Caractéristiques de l'ADN mitochondrial et son rôle en identification génétique**

L'ADNmt possède un ensemble de caractéristiques qui lui confèrent l'avantage d'être un outil innovant dans l'identification génétique.

Comparé à l'ADN nucléaire, l'ADN mitochondrial est de petite taille et est présent dans la cellule en plusieurs copies (100-1000), il est également protégé par la double membrane mitochondriale. Ces spécificités favorisent sa préservation post-mortem et sa résistance aux facteurs environnant qui attaquent et dégradent l'ADN (sol, rayons UV, humidité...etc.) et permettent son exploitation pour des échantillons anciens et dégradés. De ce fait son obtention est mieux assurée pour les analyses dans des cas où le matériel génétique disponible est de faible quantité ou dans un état très dégradé (ossement, dents, tiges de cheveux, salive, sang, etc.) (Budowle *et al.*, 2003 ; Kobilinsky *et al.*, 2005 ; Goodwin, 2007 ; Butler, 2010 ; Edlund, 2010).

L'ADNmt est le marqueur génétique de la lignée maternelle, il est strictement transmis par la lignée maternelle à travers les générations. Ce mode de transmission uniparental est dû au fait qu'un grand nombre de mitochondries paternelles soit condensé au niveau de la pièce intermédiaire du spermatozoïde alors que lors de la fécondation seule la tête du spermatozoïde, contenant un nombre très faible de mitochondries pénètre dans l'ovule qui possède des centaines de copies. Des études suggèrent qu'il existe des mécanismes de reconnaissance (notamment l'ubiquitination) qui éliminent même le petit nombre de mitochondries de provenance paternelle (Holland et Parsons, 1999 ; Sato et Sato, 2013).

Le taux de mutations de l'ADNmt qui est 5-10 fois supérieur à celui de l'ADN nucléaire. Au niveau de la région de contrôle CR, ce taux est beaucoup plus élevé et les mutations se produisent préférentiellement au niveau des sites qualifiés de régions hypervariables ou aussi de points chauds mutationnels. Ces variations génétiques héréditaires et le fait que le génome mitochondrial ne subisse pas de recombinaison permettent d'établir des liens génétiques entre les membres d'une même lignée maternelle (Budowle *et al.*, 2003 ; Kobilinski *et al.*, 2005 ; Pakendorf et Stoneking, 2005).

Grace à ces propriétés, le séquençage de l'ADNmt joue un rôle dans la résolution des enquêtes par l'identification des personnes disparues ou des victimes de guerres, de catastrophes et des tragédies, lorsque des échantillons biologiques des membres de la même lignée maternelle sont disponibles, une comparaison de l'ADNmt pourra être réalisée.

L'analyse de l'ADNmt permet dans un autre cadre, d'effectuer des études anthropologiques à partir de restes d'ossements ou de dents, ainsi que des études évolutives en retraçant les

lignages apparentés à travers le temps et en recherchant l'ascendance maternelle d'une population (Budowle et al., 2003 ; Pakendorf et Stoneking, 2005 ; Samehsalari et Reddy, 2018).

## I.4. Caractéristiques des haplotypes de l'ADN mitochondrial

### I.4.1. La séquence d'Anderson

La séquence obtenue lors du typage de la région de contrôle (CR) du génome mitochondrial est comparée à une séquence de référence, les points de différences seront retenus pour une comparaison ultérieure.

La séquence de référence est la CRS (**C**ambridge **R**eference **S**equence) aussi connue sous le nom de séquence d'Anderson, elle a été séquencée en 1981 dans le laboratoire de Sanger par Anderson et al.. Cette dernière provenait des cellules de placenta d'une femme d'ascendance européenne et a servi depuis, comme référence pour plusieurs études génétiques mais son utilisation a montré plusieurs différences.

De ce fait, la CRS a été reséquencée à partir du même échantillon utilisé par Anderson et corrigé par Andrews et ses coéquipiers en 1999. Onze erreurs frappantes ont été constatées dans la séquence originale y compris une cytosine supplémentaire en positions 3107 et des affectations incorrectes de paires de bases simples dont trois d'entre elles provenaient d'une contamination par des spécimens de bovin et de HeLa (une lignée cellulaire cancéreuse). Sept positions nucléotidiques correspondaient à un polymorphisme rare, six d'entre elles sont des substitutions d'une seule paire (au niveau des positions 263A, 750A, 1438A, 4769A, 8860A et 15326A) et la septième implique une répétition de résidus cytosine entre les positions 311 et 315. Le reste des nucléotides ainsi que la région de contrôle (HVI, HVII, HVIII) ont été confirmés identiques à la CRS originale. La CRS reséquencée porte le nom de rCRS (**r**evised **C**ambridge **R**eference **S**equence) (Anderson et al., 1981 ; Andrews et al., 1999 ; Butler, 2011 ; Bandelt et al., 2014 ; Samehsalari et Reddy, 2018).

**Tableau I.** Différences nucléotidiques entre la CRS et la rCRS

Position du nucléotide	Séquence originale	Séquence révisée	Remarque
263	A	A	Polymorphisme rare
311-315	CCCCC	CCCCC	Polymorphisme rare
750	A	A	Polymorphisme rare
1438	A	A	Polymorphisme rare
3106-3107	CC	C	Erreur
3423	G	T	Erreur
4769	A	A	Polymorphisme rare
4985	G	A	Erreur
8860	A	A	Polymorphisme rare
9559	G	C	Erreur
11335	T	C	Erreur
13702	G	C	Erreur
14199	G	T	Erreur
14272	G	C	Erreur (bovin)
14365	G	C	Erreur (bovin)
14368	G	C	Erreur
14766	T	C	Erreur (HeLa)
15326	A	A	Polymorphisme rare

(Andrews et al., 1999).



#### I.4.2. Hétéroplasmie de l'ADNmt

Les cellules d'un organisme contiennent plusieurs mitochondries chacune possédant sa propre copie d'ADNmt qui peut être différente d'une mitochondrie à une autre, donnant un mélange de plus d'un type d'ADNmt (au niveau du même tissu ou dans deux tissus différents) appelé Hétéroplasmie. Ce phénomène résulte de la transmission par la mère, de mitochondries ayant deux types d'ADNmt différents et peut être dû à une mutation spontanée ou lors de la réplication au niveau des mitochondries de l'individu. De plus certaines mitochondries de spermatozoïde peuvent échapper au mécanisme de destruction lors de la fécondation donc l'individu deviendra hétéroplasmique pour l'ADNmt maternel et paternel (Goodwin, 2007 ; Edlund, 2010 ; Ladoukakis et Zouros, 2017).

La présence de l'hétéroplasmie peut être de deux types distincts qui sont l'hétéroplasmie de longueur et hétéroplasmie de séquence.

- ❖ L'hétéroplasmie de longueur est la conséquence du glissement de l'ADN polymérase pendant la réplication des séquences répétitives, ce qui conduit à des délétions ou des additions de nucléotides au niveau de ces séquences, donc les variantes diffèrent par le nombre de bases. Ce type d'hétéroplasmie se produit aux environs de la région poly « C » (homopolymère) dans la HVI aux positions 16184 à 16193 et la HVII aux positions 303 à 315.
- ❖ Dans l'hétéroplasmie de séquence les variantes diffèrent par le type de base et due aux mutations ponctuelles de l'ADNmt. Cette catégorie d'hétéroplasmie est détectée par la présence de deux nucléotides dans une seule position (superposable) dans un électrophorogramme de séquence (Kobilinsky *et al.*, 2005 ; Butler, 2011 ; Berger *et al.*, 2011 ; Samehsalari et Reddy, 2018).

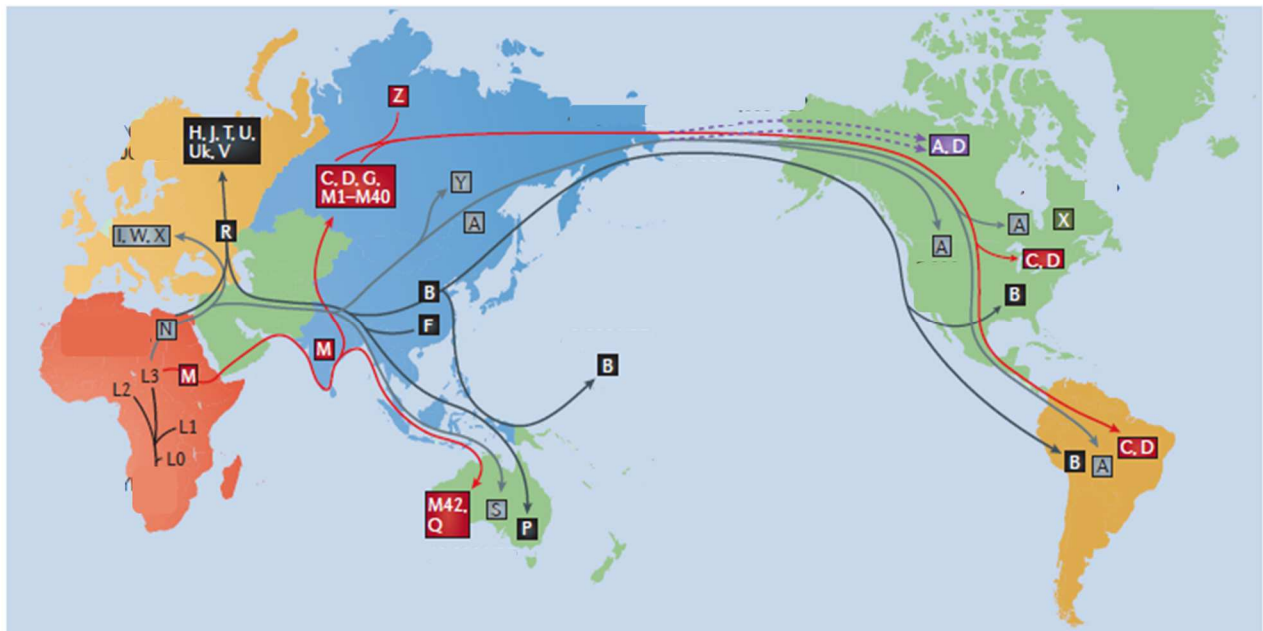
Les points chauds de l'hétéroplasmie se trouve en HVI aux positions : 16093, 16129, 16153, 16189, 16192, 16293 et 16337 ; et en HVII aux positions : 72, 152, 189, 207 et 279 (Butler, 2011).

L'hétéroplasmie a eu un rôle important en criminalistique essentiellement pour renforcer le profil génétique. Différents cas d'identification par ADNmt ont confirmé l'importance de l'hétéroplasmie tel que l'identification de la famille Romanov à partir de restes d'ossement du frère de Romanov le Tsar Nicolas II de Russie. L'Analyse de ses restes ainsi que ceux de Georgij Romanov ont montré la même hétéroplasmie à la position 16169 de la région de contrôle de l'ADNmt (Ivanov *et al.*, 1996 ; Budowle *et al.*, 2003 ; Edlund, 2010).

#### I.4.3. Les haplogroupes et la base de données EMPOP

Le polymorphisme de l'ADNmt est représenté par des insertion, des délétions ou des différences au niveau d'une seule base (SNP). Ces variations sont héritées de génération en génération à travers la lignée maternelle, ce qui en résulte un haplotype qui est donc un ensemble de polymorphisme situé sur la même région de l'ADNmt (CR : HVI, HVII, HVIII), ainsi plusieurs haplotypes partageant les mêmes variations peuvent constituer un haplogroupe (Budowle *et al.*, 2003 ; Lembring, 2013).

Plusieurs études phylogénétiques ont permis de mettre en évidence différents haplogroupes selon la zone géographique. Il a été constaté que le niveau de diversité est plus haut chez les Africains alors que les Amérindiens enregistrent le plus bas. L'origine de la phylogénie de l'ADNmt et les branches les plus diverse sont également, limités aux populations africaines (Wallace, 1999 ; Kivisild, 2015). Les haplogroupes A, B, C, D sont associés aux populations Amérindiennes (Van Oven et Kayser, 2008 ; Kivisild, 2015). Alors que la lignée L (L0 à L6) est fréquente dans les régions africaines sub-sahariennes, avec les haplogroupes L0, L1, L2 et L3 spécifique de la population africaine et U6 étant plus fréquent dans les populations nord-africaines. Les lignées fréquentes en sud de l'Asie sont M et N (dérivée de la lignée L3), A, B, C, D, E, F, G, P, Q, R, Y, et Z. Ainsi les populations européennes ont des fréquences élevées d'haplogroupes H, I, J, K, T, U, V, W et X (Umetsu et Yusua, 2005 ; Courday *et al.*, 2009 ; Kivisild, 2015). Ces haplogroupes peuvent être retrouvés dans des régions autres que les régions d'origine, mais avec une référence rare et cela dû aux flux migratoires des individus surtout celui de la femme (Kivisild, 2015) (Figure 3).



**Figure 3.** Les haplogroupes Majeurs de la population Humaine (Stewart et Chinnery, 2015)

La racine africaine était à l'origine de quatre lignées spécifiques à l'Afrique subsaharienne : L0, L1, L2 et L3. Les haplogroupes, M et N, sont issus de l'haplogroupe africain L3. Lors de la migration, l'haplogroupe N a été dirigé vers l'Eurasie et les lignées d'haplogroupe M ont été déplacées en Asie, donnant naissance aux haplogroupes A, B, C, D, G et F. En Europe, l'haplogroupe N a conduit à l'haplogroupe R, qui est la racine des haplogroupes européens H, J, T, U et V. Les haplogroupes S, P et Q se trouvent en Australasie et les haplogroupes A, B, C et D sont apparus et peuplés en Asie de l'Est et dans les Amériques.

La classification des haplogroupes est essentiel en identification génétique, ce qui rend l'analyse d'ADNmt plus discriminante en comparant le profil obtenu aux bases de données des haplogroupes et ainsi identifier la zone géographique de l'individu (Reid, 2018).

Les différents haplogroupes sont répertoriés dans l'EMPOP « European DNA Profiling Group (EDNAP) mtDNA Population Database » qui est une base de données Web, conçue pour la gestion des données de séquences d'ADNmt de diverses populations dans le monde et pour l'évaluation de la rareté d'un haplotype d'ADNmt dans une population donnée. Elle représente un projet collaboratif regroupant des laboratoires de criminalistique et de génétique de population dans le monde entier dans le but d'augmenter la quantité et la qualité de données fiables sur les séquences d'ADNmt dans un format consultable via internet sur le site « <https://empop.online> ». Les haplotypes ont fait l'objet d'une révision minutieuse à l'aide des logiciels qui ont été installés dans la base de données afin de faire une inspection des données et de faire face aux taux élevés d'erreurs d'interprétation des données brutes et/ou l'introduction d'erreurs d'écriture dans le processus de transcription des données (Parson *et al.*, 2004<sup>a</sup> ; Parson *et al.*, 2004<sup>b</sup> ; Parson et Dür, 2007 ; Prieto *et al.*, 2011 ; Huber *et al.*, 2018).

## I.5. Validation des méthodes d'analyse

### I.5.1. Définition de la validation

La validation est un processus comportant un ensemble d'opérations, permettant d'obtenir les informations nécessaires à l'évaluation de l'efficacité et de la fiabilité d'une procédure, d'un protocole, d'une expérience ou d'une méthode de laboratoire. La validation permet en outre de vérifier la capacité d'une méthode à obtenir un résultat souhaité pour un usage déterminé (Blanchin, 2010 ; Butler, 2011 ; SWGDAW, 2016 ; ASB Standard, 2018).

On distingue deux types de validation, la validation de développement et la validation interne :

#### ➤ Validation de développement

Cette validation est effectuée par le fabricant du kit, du test, ou de l'outil utilisé dans la méthode. Elle consiste à acquérir des données d'essai et à déterminer les conditions et les limites d'une nouvelle méthode d'analyse des échantillons (Blanchin, 2010 ; Butler, 2011 ; SWGDAW, 2016 ; ASB Standard, 2018).

#### ➤ Validation interne

Ce type de validation est réalisé par un laboratoire particulier venant d'introduire une nouvelle méthode. Elle implique la vérification et la démonstration que les méthodes, les résultats et l'interprétation des données se déroulent efficacement et dans les limites prédéterminées dans le laboratoire (Blanchin, 2010 ; Butler, 2011 ; SWGDAW, 2016 ; ASB Standard, 2018).

### I.5.2. Objectifs de la validation

En raison de leur rôle dans le système de justice pénale, et en particulier dans l'identification des personnes par l'analyse de l'ADN, les laboratoires forensiques sont tenus de faire une validation de leurs méthodes afin de répondre à la norme internationale « ISO 17025 » qui définit les exigences pour l'accréditation des laboratoires dans le cadre de l'assurance de la qualité. En effet, une méthode validée et normalisée est mieux acceptée par la justice, en plus de la possibilité de comparer la technique au niveau international entre différents laboratoires. La validation a pour objectif (i) d'obtenir des résultats d'une méthode, avec une qualité de mesure pertinente (ii) que ces résultats soient cohérents, fiables et précis (iii) définir

rigoureusement les conditions requises pour avoir précisément ces résultats, (iv) de déterminer les limites de la procédure et donc de certifier que son résultat et son usage sont comme envisagés. Ce processus de validation permet aussi d'identifier les aspects de l'analyse (variations de ses paramètres), qui doivent être surveillés et contrôlés, et, sert de base à l'élaboration de directives d'interprétation afin de communiquer l'importance des résultats (Budowle *et al.*, 2008 ; Blanchin, 2010 ; Butler, 2014 ; ASB Standard, 2018).

### I.5.3. Critères de la validation

Les critères de la procédure de validation sont divers (exactitude, fidélité, linéarité, reproductibilité, répétabilité, spécificité, sensibilité, robustesse, sélectivité, etc.), et leur étude est essentielle pour la validation de la méthode analytique.

Dans cette étude de validation interne du protocole de séquençage des régions hypervariables de l'ADNmt pour des échantillons de référence, les critères retenus sont :

- **Exactitude (justesse) :** est la capacité, pour une analyse, de fournir les mêmes valeurs d'un résultat ou une valeur suffisamment proche, à celle acceptée et conventionnelle qui est une valeur de référence. Autrement dit, ce critère mesure l'étroitesse d'accord entre la valeur de l'analyse et la valeur réelle (Blanchin, 2010 ; SWGDAW, 2016).
- **La Fidélité :** représente l'ensemble des caractéristiques de dispersion. On distingue :
  - La répétabilité :** est la capacité à reproduire l'analyse du même échantillon à partir de prises d'essais multiples avec les conditions inchangées, ainsi ce critère mesure la différence maximale acceptable entre une série de résultats obtenus simultanément par le même manipulateur et sur le même échantillon (Hlinka *et al.*, 2009 ; Blanchin, 2010 ; SWGDAW, 2016).
  - La reproductibilité :** est une mesure de la différence entre deux résultats d'une analyse obtenus par différents opérateurs et à des temps différents mais sur le même échantillon (Hlinka *et al.*, 2009 ; Blanchin, 2010 ; SWGDAW, 2016)

# **Chapitre II : Matériel et Méthodes**

## Chapitre II : Matériels et méthodes

Cette étude a été réalisée au niveau du laboratoire du Département de Biologie de l'Institut National de Criminalistique et de Criminologie de la Gendarmerie Nationale (INCC-GN) d'Alger.

L'objectif de ce travail est la contribution à l'élaboration et la vérification d'un protocole d'analyse de « l'ADNmt de source connue » dans le but de l'identification génétique.

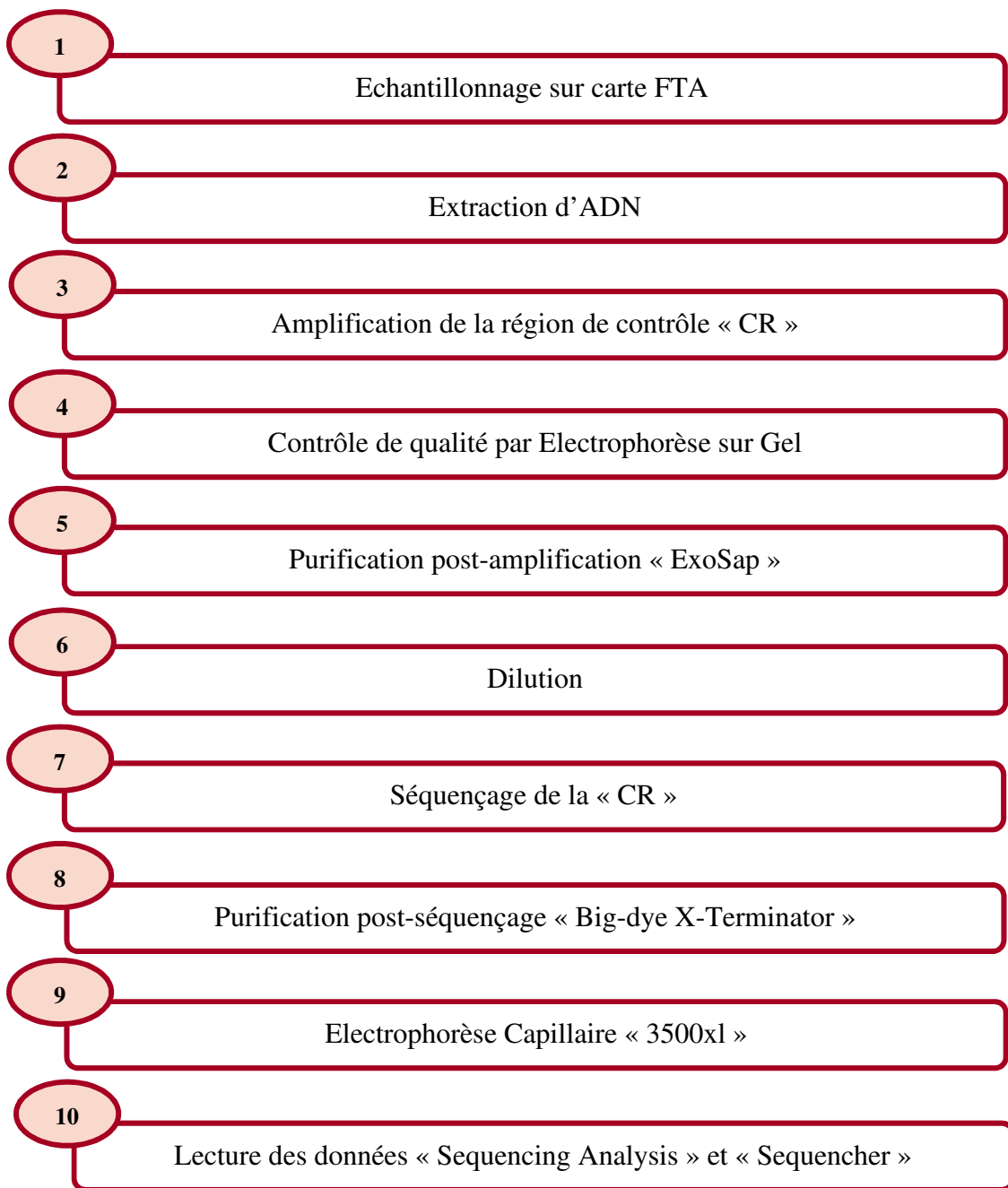
Ce protocole serait avantageux dans le cas où l'échantillon présent serait un matériel biologique dont l'ADNnuc n'est pas accessible (dégradés ou en faible quantité) et que l'ADNmt est une séquence intacte et continue, ainsi que dans le cas d'identification d'individus en se basant sur la lignée maternelle.

Les échantillons utilisés pour cette étude sont de type salive obtenus à partir d'un prélèvement buccal, ainsi que trois contrôles d'ADN connus :

- ❖ Le 2800M fournis par la firme Promega
- ❖ Deux standards CHR et GM09947A du SRM2392 fournis et séquencés par le NIST (National Institute of Standards and Technology).

Ces échantillons ont été utilisés pour le séquençage des trois régions hypervariables de la région de contrôle de l'ADNmt (HVI, HVII, HVIII). La méthodologie a été effectuée suivant trois stratégies différentes comprenant 9, 10 ou 11 amorces (primers).

Le protocole débute par l'échantillonnage sur carte FTA, suivie par la mise en tube des punches. L'étape suivante est l'amplification par PCR de la région de contrôle « CR » (HVI, HVII, HVIII). Ensuite dans le but d'une vérification du bon déroulement de l'amplification, une électrophorèse sur Gel est effectuée. Après confirmation et succès de la 3<sup>ème</sup> étape, l'ADN amplifié subit une purification post-amplification et une dilution. Par la suite l'ADN amplifié et purifié est utilisé pour une 7<sup>ème</sup> étape qui est le séquençage des régions hypervariables de la « CR », par PCR et en utilisant la stratégie des amorces adéquates. Dès lors, une seconde purification post-séquençage est effectuée sur l'ADNmt séquencé. Le résultat des séquençages obtenus est recueilli suite à la migration par électrophorèse capillaire. Enfin, une dernière étape consiste en la lecture des données par deux logiciels spécialisés afin de visualiser les données brutes et de réaliser un alignement des séquences obtenues avec la « rCRS » et les différences remarquées constituent le profil de chaque échantillon (**Figure 4**).



**Figure 4.** Diagramme représentatif des étapes du protocole d'analyse de l'ADNmt

## II.1. Recommandation pour une analyse de l'ADNmt Référence

L'analyse de l'ADNmt est très sensible et demande des mesures et des précautions strictes et précises.

Tout d'abord, le manipulateur doit porter les équipements de protections et d'hygiène (Gants, Masque d'Hygiène, Charlotte).

Avant chaque manipulation d'une étape du protocole la paillasse doit être nettoyée à l'eau de Javel ou le nettoyeur RBS et aussi stérilisée à la lampe UV de même pour le matériel utilisé (Micropipette, Micro-tubes, Tubes, Portoirs, etc.).

Chaque étape doit être manipulée et réalisée sous hotte à flux laminaire ou sous hotte BioCap. L'ADNmt est traité dans une salle différente et séparée de celle où la préparation du mélange réactionnel est effectuée.

Concernant ce protocole d'analyse de l'ADNmt Référence, l'utilisation d'un contrôle négatif pour les étapes d'extraction, d'amplification et de séquençage, est nécessaire dans le but de vérifier la présence d'une éventuelle contamination.

Un contrôle positif est recommandé pour la réaction d'amplification et de séquençage pour vérifier le bon fonctionnement des réactifs.

Lors de l'étape de séquençage les amorces Forward et Reverse « F et R » sont traitées et utilisées séparément afin d'éviter toute contamination entre les deux amorces et ainsi ne pas altérer les résultats de séquençage avec une double lecture en simultané.

Enfin, en terminant chaque étape de manipulation un nouveau nettoyage est nécessaire.

## II.2 Echantillonnage

Pour l'étude de ce protocole, les échantillons sont obtenus à partir d'un prélèvement de salive par un frottis buccal de la face interne de la joue à l'aide d'un écouvillon buccal stérile et déposés sur une carte FTA. 13 cartes FTA prélevées d'individus anonymes et consentants, préalablement préparées et référencées portant un code d'identification E9, E10, E17, E18, E115, E151, E172, E195, E215, E231, E234, E630 et E631, nous ont été fournies par le laboratoire. 3 ADN contrôles standard dont le profil génétique mitochondrial (mitotype) est connu : 2800M (M<sup>+</sup>), CHR (C<sup>+</sup>) et GM09947A (T<sup>+</sup>) nous ont été également fournis. De ce fait nous avons utilisées 16 échantillons au total.

## II.3. Extraction

### II.3.1. Principe de la FTA

La FTA (Whatman™ FTA™) est une carte contenant une matrice imprégnée d'une formule chimique brevetée. Cette chimie permet la lyse des cellules, la dénaturation des protéines et l'élimination des contaminants en déposant l'échantillon biologique. Tandis que les acides nucléiques sont enchâssés et protégés dans les fibres de la matrice. La carte FTA peut être



transportée au laboratoire et stockée à température ambiante en vue d'une analyse ultérieure (Parsons et Bright, 2012).

### II.3.2. Poinçonnage et lavage

Pour récupérer l'ADN purifié qui se trouve sur les deux cercles de la carte FTA, une étape de poinçonnage et une série de lavages sont nécessaires. Dans un premier lieu, 50 µl du tampon TE (Tris-EDTA) ont été déposés, dans des micro-tubes Eppendorf de 0.2 ml, préalablement identifiés selon le code de l'échantillon. Après avoir déposé la carte sur un tapis de découpe (Harris cutting mat), deux poinçonnages ont été effectués, en utilisant un stylo micro emporte-pièce (Harris Uni-Core punch), sur les treize échantillons respectivement puis sur une carte FTA vide ne contenant aucun prélèvement qui sert de contrôle de contamination (contrôle négatif). Les poinçonnages ont été mis en micro-tubes contenant le tampon de solubilisation TE. Ensuite les 50µl de ce dernier ont été aspirés et éliminés.

Pour un premier lavage, 100 µl du TE ont été ajoutés dans chaque micro-tube. Sur un portoir, les micro-tubes sont passés à l'agitateur pendant 3 minutes, suivi d'une centrifugation rapide (quick spin) ; ensuite, les 100 µl du TE ont été aspirés et éliminés. Le deuxième lavage a été effectué avec les mêmes étapes sauf que, l'agitation a été faite pendant 2 minutes et 30 secondes.

## II.4. Amplification par PCR

### II.4.1. Principe

PCR « Polymerase Chain Reaction » est une technique enzymatique permettant d'amplifier une séquence d'ADN, c'est-à-dire fabriquer en grande quantité un même fragment d'ADN. Chaque analyse PCR nécessite la présence d'ADN matrice, d'amorces, de nucléotides et d'ADN polymérase. Les amorces sont de courts fragments d'ADN avec une séquence définie complémentaire à l'ADN cible à détecter et à amplifier. Celles-ci servent de point d'extension à l'ADN polymérase (Garibyan et Avashia, 2013).

La PCR se base sur les caractéristiques d'hybridation de l'ADN en fonction de la température (dénaturation et hybridation), et de polymérisation par une enzyme thermostable. Premièrement, l'ADN est dénaturé à haute température (95°C) ; ensuite, cette dernière est abaissée rapidement afin de permettre l'hybridation des amorces. Cette étape est cruciale et la température d'hybridation des amorces est spécifique, elle dépend de la nature et de la longueur des amorces. Enfin, l'élongation des amorces est effectuée par la Taq polymérase qui utilise les dNTPs du milieu et nécessite du  $Mg^{2+}$ , cette dernière étape se réalise après que la température s'élève à 72°C. Le produit final est ensuite à nouveau dénaturé pendant un autre cycle et après ces trois étapes, le cycle est répété de 20 à 60 fois (Bell, 1989 ; Ishmael et al., 2008) (Figure 5).

### II.4.2. Choix des amorces

Ce travail se base sur le séquençage des brins sens et antisens des trois régions hypervariables (HVI, HVII, HVIII) de la CR, permettant ainsi d'utiliser la CR en entier comme séquence d'intérêt pour l'amplification (16024 à 576). Pour cela, deux types d'amorces ont été

choisis : le F15971 pour amorcer le brin sens et le R614 pour amorcer le brin antisens (**Tableau II**).

Le choix de ces deux amorces a été effectué en prenant en considération le fait que nous avons travaillé avec une séquence complète et intacte (non dégradée) puisque les échantillons utilisés pour cette analyse sont de l'ADNmt référence et particulièrement sa région de contrôle (CR). Ainsi, pour éviter les erreurs de lecture, ces deux amorces s'hybrident un peu avant le début de la séquence d'intérêt.

**Tableau II.** Les amorces d'amplification utilisés

Amorces	Séquence	Région	Taille	Brin
<b>F15971</b>	5'-TTAACTCCACCATTAGCACC-3'	HVI	20 nucléotides	Sens
<b>R614</b>	5'-TTTCAGTGTATTGCTTTGAGGA-3'	HVIII	22 nucléotides	Antisens

#### II.4.3. Amplification par le kit « AmpliTaq Gold 360 Master Mix »

La réaction de PCR a été effectuée en utilisant un seul mélange réactionnel d'un volume réactionnel total de 25  $\mu$ l pour chaque échantillon contenant : 9.5  $\mu$ l d'eau ultrapure, 1  $\mu$ l amorce F15971, 1  $\mu$ l amorce R614 (les deux amorces s'utilisent simultanément) et 12.5  $\mu$ l du réactif « AmpliTaq Gold 360 Master Mix », conçu pour amplifier l'ADNmt et contient les éléments nécessaires pour une PCR : la Taq polymérase et les dNTPs. Sur lesquels, a été rajouté 1  $\mu$ l de BSA « Bovin Serum Albumin » qui joue le rôle d'un co-amplificateur en augmentant le rendement de la PCR.

Avant de procéder à la préparation du mélange réactionnel, il faut s'assurer que la concentration des amorces est de 10 pmol/ $\mu$ l (10  $\mu$ M) et que celle de la BSA est de 0.25  $\mu$ g/ $\mu$ l.

Pour commencer la réaction d'amplification, un mélange réactionnel (AmpliTaQ Golg 360 Master Mix + BSA + amorce F + amorce R + eau ultrapure) a été préparé pour dix-sept échantillons plus un, comme indiqué dans le **tableau III**.

**Tableau III.** Calculs des volumes du mélange réactionnel de l'amplification

		Quantité par nombre d'échantillons	
		1 échantillon	17 échantillons + 1
Composants du mélange réactionnel	AmpliTaQ Gold 360 Master Mix	12.5 $\mu$ l	225 $\mu$ l
	BSA	1 $\mu$ l	18 $\mu$ l
	Amorce F15971	1 $\mu$ l	18 $\mu$ l
	Amorce R614	1 $\mu$ l	18 $\mu$ l
	Eau ultrapure	9.5 $\mu$ l	171 $\mu$ l
	Volume Total	25 $\mu$ l	450 $\mu$ l

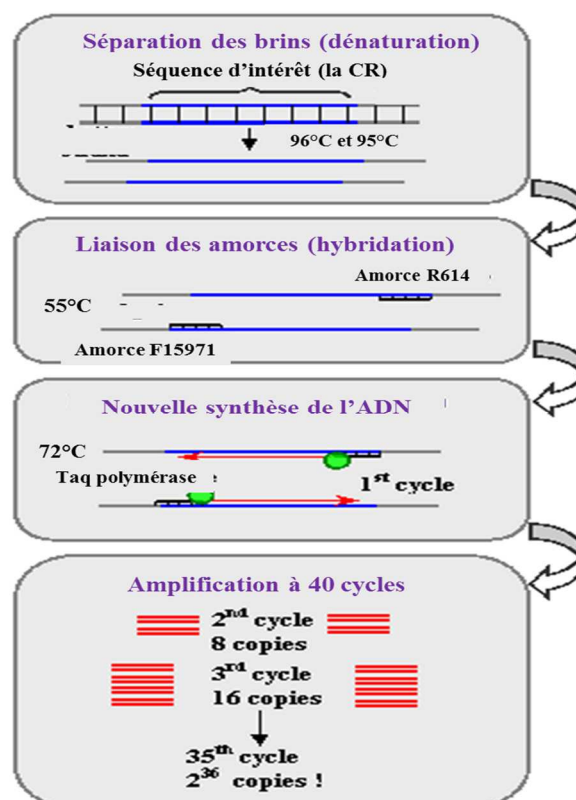
Dans un tube de 1.5ml stérilisé et nommé « Mix Amplif », le volume total 450µl du mélange réactionnel a été déposé comme suit : 225 µl d'AmpliTaQ Gold 360 Master Mix ont été ajoutés, suivis de 171 µl d'eau ultrapure, 18 µl de BSA, 18 µl d'amorce F et 18 µl d'amorceR. Après, le mélange passe au vortex pour quelques secondes afin d'homogénéiser le tout. Ensuite, 25 µl de ce mélange ont été prises pour chaque échantillon et déposés dans les 14 micro-tubes préparés précédemment contenant les poinçonnages (13 ADN et un témoin négatif).

Pour les 3 témoins contrôles, 1 µl de chaque aliquot (l'ADN témoin dilué selon les recommandations du fabricant pour obtenir une concentration optimale de 1ng/µl) a été prélevé et déposé dans trois micro-tubes de 0.2ml stérilisés, nommés selon le témoin : T<sup>+</sup>, C<sup>+</sup> et M<sup>+</sup> ; puis, le volume a été complété par 24 µl du mélange réactionnel.

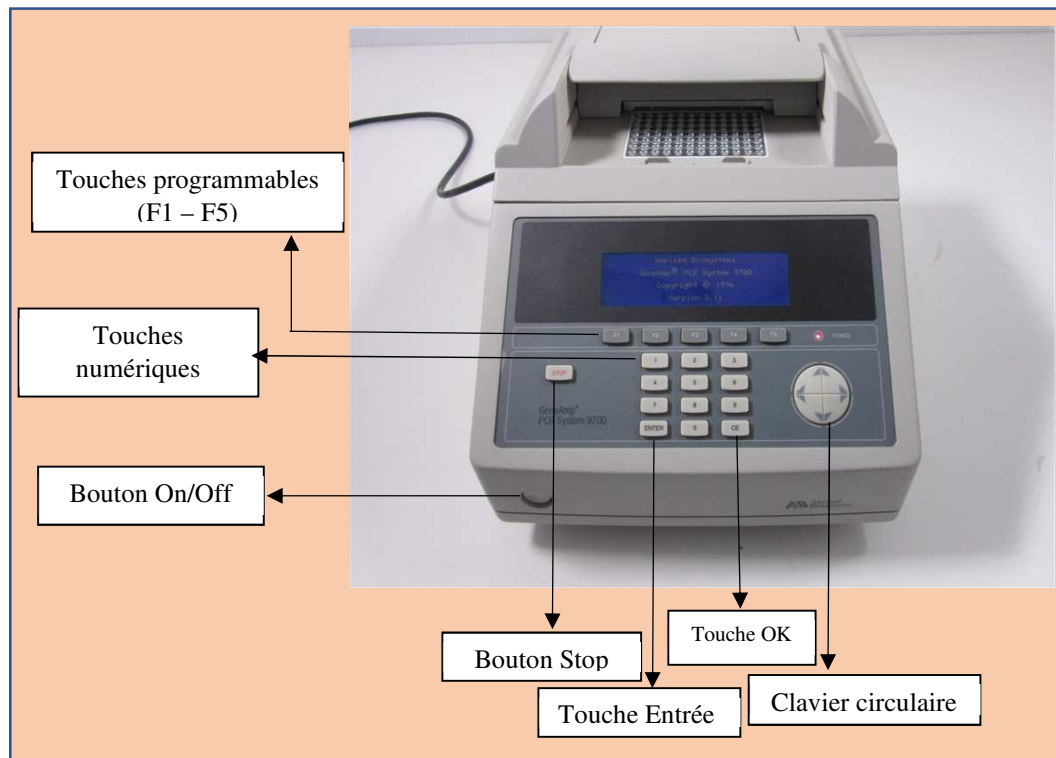
Les échantillons ont été placés dans le thermocycleur « Gene Amp PCR System 9700 - Applied biosystems » (**Figure 6**) pour réaliser une PCR à 40 cycles selon le programme suivant : (**Tableau IV**)

**Tableau IV.** Programme thermocycleur pour l'amplification

Etape de dénaturation initiale	40 Cycles			Extension (élongation) finale	Etape de maintenance
	Dénaturation	Hybridation	Extension (élongation)		
Maintien	Cycle			Maintien	Maintien
96°C, 10 minutes	95°C, 30 secondes	55°C, 1 minute	72°C, 60 secondes	72°C, 7 minutes	4°C, ∞



**Figure 5.** Les étapes de la PCR (NCBI, 2017)



**Figure 6.** Thermocycleur Gene Amp PCR System 9700 - Applied biosystems (Marshall scientific)

## II.5. Contrôle de qualité par électrophorèse sur gel

Après avoir amplifié la CR de l'ADNmt, les amplicons passent par une électrophorèse sur gel d'agarose, qui permet à la fois de vérifier le bon déroulement de la réaction d'amplification et d'estimer approximativement la quantité d'ADNmt pour chaque échantillon.

### II.5.1. Principe

L'électrophorèse sur gel d'agarose est une méthode relativement rapide et facile pour évaluer à la fois la qualité et la quantité de l'ADN extrait. Elle se base sur la séparation des fragments d'ADN selon leur taille et leur charge négative par migration en appliquant un champ électrique. Le gel est immergé dans un tampon d'électrophorèse et l'ADN est chargé dans des puits formés dans le gel par un peigne. Le gel d'agarose forme une matrice poreuse et agira comme un tamis lors de la procédure d'électrophorèse donc les plus petites molécules d'ADN traversent le gel plus rapidement que les plus grandes molécules d'ADN. Des colorants qui s'intercalent avec la double hélice de l'ADN, s'ajoutent avant ou après l'électrophorèse. L'ADN est visualisé en plaçant le gel sur un transilluminateur émettant une lumière UV. Des standards de quantification sont ajoutés à côté des échantillons pour permettre l'estimation de la taille et la quantité d'ADN. (Kobilinsky *et al.*, 2005 ; Goodwin *et al.*, 2007)

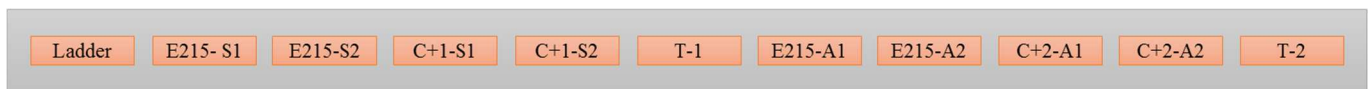
### II.5.2. Préparation du gel pour 180ml

Un gel à 2% d'agarose a été préparé en diluant la poudre de ce dernier dans une solution tamponnée de TAE dilué à 1X. Ce mélange a été placé au microonde pour homogénéisation pendant 2minutes en vérifiant et en mélangeant toutes les 30 secondes. Dès que la solution est

devenue transparente et homogène, le colorant SYBR SAFE a été rajouté à un volume constituant 10% du gel. Ensuite, le contenu a été versé délicatement dans le moule et le peigne a été installé sur le montage ; le gel est laissé au frais pour se solidifier pendant 30 minutes au minimum. Après, le peigne est enlevé du gel délicatement, le moule de ce dernier est placé dans le système de migration et couvert totalement avec le TAE restant de la préparation 1X.

### II.5.3. Dépôt et migration sur gel

Dans des tubes de 0.2 ml stérilisés et identifiés selon l'échantillon, 4 µl de chaque produit amplifié et 1 µl de Blue Juice (colorant) ont été déposés et bien mélangés à la pipette. 5µl du Low DNA Mass Ladder ont été placés dans le premier puits, suivis par le volume préparé dans chaque puits en respectant un plan de dépôt indiquant la position de chaque échantillon (**Figure 7**). Le montage a été fermé en respectant la position de la cathode - (noire) et de l'anode + (rouge). Le niveau du TAE a été vérifié. Le générateur de courant a été allumé et les paramètres de migration ont été réglés à : courant 150V. La migration se fait pendant 30 minutes.



**Figure 7.** Exemple d'un plan de dépôt sur gel utilisé

### II.5.4. Visualisation des bandes et estimation de la quantité

Le moule a été retiré et le gel est glissé sur la plaque de visualisation UV ; cette dernière est recouverte avec le couvercle en verre et allumée.

## II.6. Purification post-amplification

### II.6.1. Principe

La purification post-amplification est une étape importante pour nettoyer les amorces et les dNTPs non incorporés qui peuvent interférer avec les réactions ultérieures et conduire à une séquence illisible. Les méthodes classiques utilisées pour purifier les produits de PCR avant le séquençage comprennent l'électrophorèse sur gel, la précipitation à l'éthanol et la chromatographie sur colonne. Toutefois, ces méthodes peuvent présenter divers degrés de perte d'échantillon, alors que, la méthode enzymatique par ExoSAP-IT® permet d'éviter ce problème en plus d'être simple à exécuter et de fournir des résultats cohérents. ExoSAP-IT® est composé de 2 enzymes hydrolytiques, « Exonuclease I » et « Shrimp Alkaline Phosphatase » ; l'Exonucléase I dégrade les amorces monocaténaire résiduelles et tout ADN monocaténaire produit par la PCR, alors que la « Shrimp Alkaline Phosphatase » hydrolyse les dNTPs restants du mélange de la PCR (USB corporation, 2000 ; Bell, 2008).

### II.6.2. Purification à l'ExoSAP-IT®

Après électrophorèse sur gel 16 (13 FTA, 3 ADN+, 1 T<sup>-</sup>) échantillons ont été retenus pour poursuivre le reste du protocole. La purification par l'ExoSAP® se déroule comme suit : 5µl de chaque produit amplifié et 2µl d'ExoSAP® ont été prises et déposés dans des micro-tubes de 0.2ml nommés selon l'échantillon avec ajout de la lettre « P » pour purification. Les

échantillons sont placés sur la plaque du thermocycleur et la réaction a été opérée selon le programme suivant : 37°C pendant 15 minutes et 80°C pendant 15 minutes.

## II.7. Dilution

En se basant sur les résultats de l'électrophorèse sur gel, les produits amplifiés, quantifiés et purifiés ont été dans un premier temps dilués. Les échantillons ont été dilués 1/6<sup>ème</sup>, en commençant par une concentration de départ de 30ng/μl pour arriver à une concentration minimale de 5ng/μl pour procéder au séquençage.

## II.8. Réaction de Séquençage

### II.8.1. Principe de la méthode

Pour le protocole étudié, la technique de séquençage utilisée a été la méthode enzymatique de Sanger, dont le principe se base sur l'utilisation d'amorces (primers), de désoxyribonucléotides « dNTPs », de didésoxyribonucléotides « ddNTPs » (constituant clé de la méthode), et de l'utilisation de l'ADN Polymérase afin d'aboutir à la fabrication d'une séquence d'ADN décryptée et dont l'ordre et l'enchaînement des nucléotides sont déterminés (Elrod et Stansfield, 2003).

La réaction de séquençage de Sanger débute par une étape de dénaturation de l'ADN cible, suivie par une hybridation spécifique des amorces à l'ADN simple brin. A partir de l'extrémité 3' des amorces, l'ADN polymérase synthétise le brin complémentaire par ajout complémentaire et aléatoire des dNTPs ou des ddNTPs. Au moment où un ddNTP est utilisé, la polymérisation est bloquée du fait de l'absence d'une extrémité 3' OH libre. L'utilisation des ddNTPs est crucial, de ce fait, ils sont marqués par des fluorophores différents (4 couleurs différentes). Au terme de cette réaction, un lot de brins d'ADN de différentes tailles est obtenu, qui seront par la suite séparés sur un gel d'électrophorèse, afin de pouvoir les identifier par fluorescence (Turner et al., 2006).

Par la suite cette technique a été automatisée ce qui l'a rendue plus facile et plus rapide à réaliser où la lecture des résultats se fait par ordinateur ce qui est plus précis (Boujard et al., 2015).

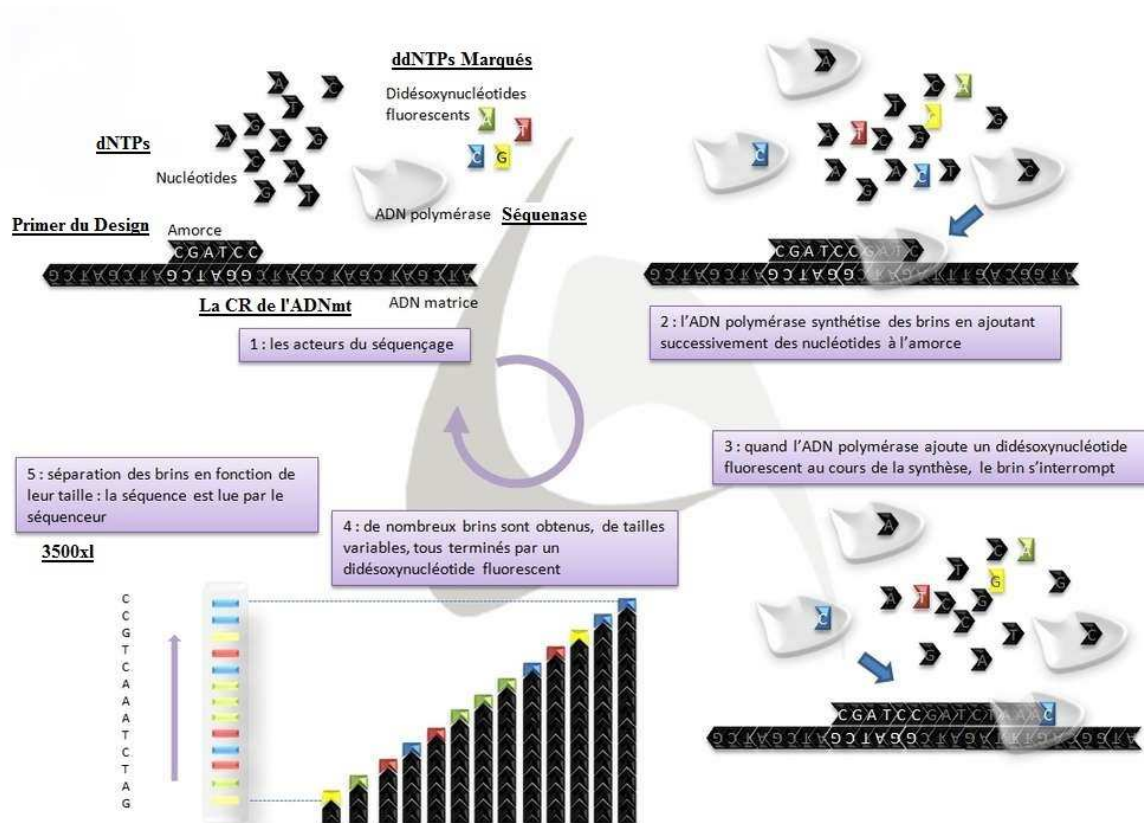


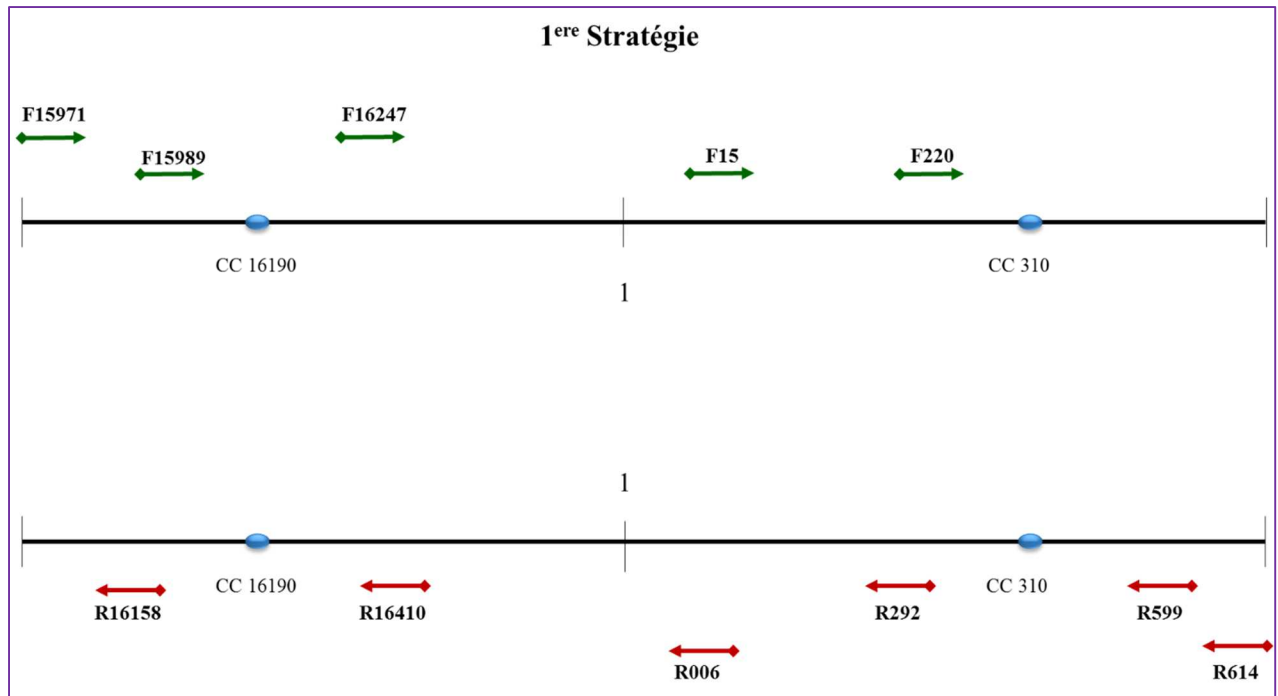
Figure 8. La réaction de séquençage de Sanger (Claire Peltier, Futura-Sciences, 2011).

## II.8.2. Mode opératoire

### a) Choix des amorces pour le séquençage

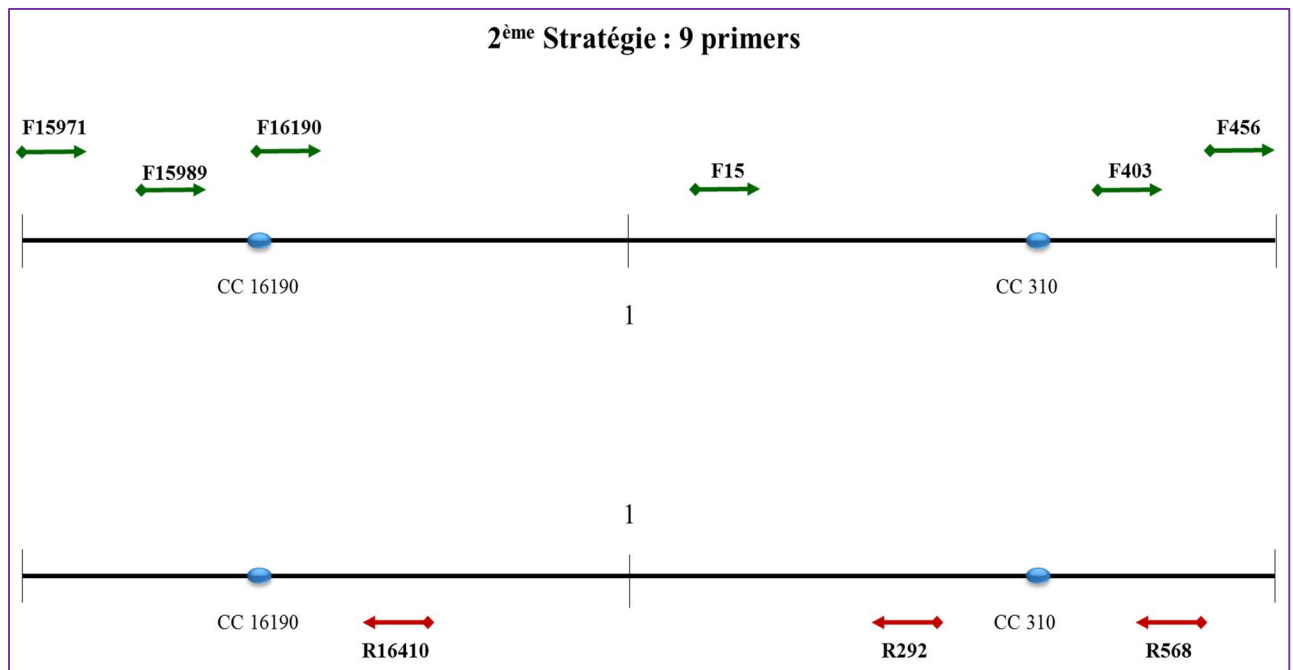
Pour le choix de nos amorces, nous avons testé quatre stratégies, dans le but de garder celui où ceux qui permettront de couvrir un maximum de surface de lecture avec un minimum de nombre d'amorces. Les amorces utilisées ont été sélectionné en se référant à d'autres travaux ainsi qu'à des articles.

1<sup>ère</sup> stratégie : 11 amorces ont été choisis « 5 Forward et 6 Reverse : F15971, F15989, F16247, F15 et F220 ; R614, R599, R292, R006, R16410 et R16158 », couvrant ainsi une lecture de 300pb entre chaque amorce (Figure 9).



**Figure 9.** Primers de la 1<sup>ère</sup> stratégie et leurs positions

Le résultat de ce séquençage montre que les primers couvraient finalement jusqu'à 400pb, nous permettant ainsi de réduire le nombre d'amorces à 9 : 6 Forward et 3 Reverse (F15971, F15989, F16190, F15, F403 et F456 ; R568, R292 et R16410) ce qui constitue la 2<sup>ème</sup> stratégie (Figure 10).



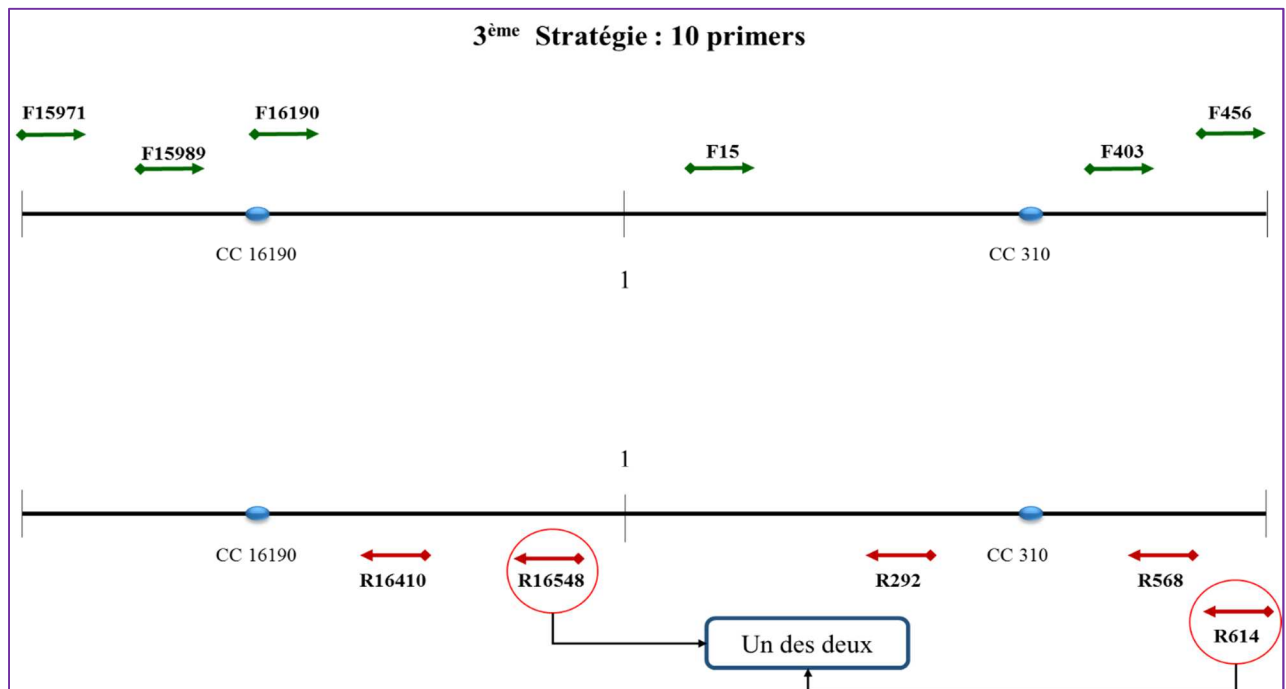
**Figure 10.** Primers de la 2<sup>ème</sup> stratégie et leurs positions.



La 3<sup>ème</sup> et 4<sup>ème</sup> stratégie, ont été proposés pour pallier au problème de la présence de la séquence « Poly C », soit au niveau de l'une des régions hypervariables : « HVI » ou « HVII » ou au niveau de ces deux régions, et sont constitués respectivement de 10 primers (6 Forward et 4 Reverse) et de 11 primers (6 Forward et 5 Reverse) (**Figure 11** et **Figure 12**).

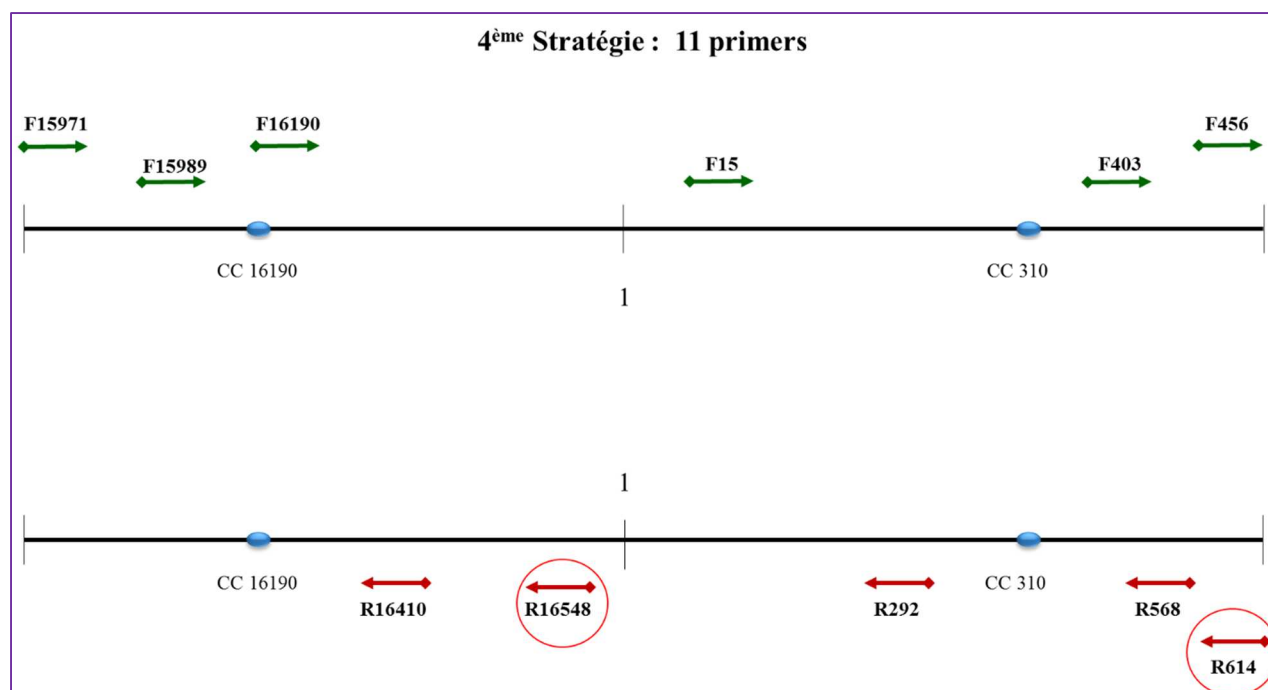
Après avoir testé les quatre stratégies, nous avons éliminé la 1<sup>ère</sup>, les trois autres stratégies ont été utilisées comme suit :

- 2<sup>ème</sup> stratégie « 9 primers » : il est utilisé dans le cas où l'échantillon à analyser contient un « Poly C » assez court en longueur.
- 3<sup>ème</sup> stratégie « 10 primers » : si l'échantillon étudié est constitué d'un « Poly C » de longueur allongée, au niveau de la HVI les primers (F15971, F15989, F16190, F15, F403 et F456 ; R568, R292, R16548 et R16410) seront utilisés. Alors que dans le cas où le « Poly C » est présent au niveau de la HVII les amorces (F15971, F15989, F16190, F15, F403 et F456 ; R614, R568, R292 et R16410) seront choisies.



**Figure 11.** Primers de la 3<sup>ème</sup> stratégie et leurs positions.

- 4<sup>ème</sup> stratégie « 11 primers » : dans le cas où l'échantillon contiendraient un « Poly C » allongé, dans les deux régions HVI et HVII, les primers choisis seront (F15971, F15989, F16190, F15, F403 et F456 ; R614, R568, R292, R16548 et R16410).



**Figure 12.** Primers de la 4<sup>ème</sup> stratégie et leurs positions.

### b) Séquençage

Après avoir mis en place le design adéquat, le mélange de la réaction de séquençage d'un volume total de 20  $\mu\text{l}$ , a été préparé et annoté en fonction du primer utilisé (par ex : Mix F15971). Dans chaque Eppendorf de 0.2 ml déposer : 13  $\mu\text{l}$  d'eau ultra-pure, 4  $\mu\text{l}$  du Sequencing Buffer, 1  $\mu\text{l}$  du primer (F ou R), 1  $\mu\text{l}$  du Ready Reaction Mix (contenant les composants essentiels au séquençage de Sanger : ddNTPs, dNTPs et l'ADN pol) et 1  $\mu\text{l}$  d'ADNmt amplifié. Pour le Témoin Négatif, dans un Eppendorf de 0.2 ml noté selon le primer (par ex : T<sup>-</sup> R292) déposer : 13  $\mu\text{l}$  d'eau ultra-pure, 4  $\mu\text{l}$  du Sequencing Buffer, 1  $\mu\text{l}$  du primer (soit F ou R), 1  $\mu\text{l}$  du Ready Reaction Mix et 1  $\mu\text{l}$  d'eau ultra-pure (**Tableau V**).

**Tableau V.** Calculs des volumes du mélange réactionnel de la réaction de séquençage

		Quantité par nombre d'échantillons	
		1 échantillon	17 échantillons + 1
Composants du Mélange Réactionnel	Ready Reaction Mix	1 $\mu\text{l}$	17 $\mu\text{l}$
	Sequencing Buffer	4 $\mu\text{l}$	68 $\mu\text{l}$
	Primer (F ou R)	1 $\mu\text{l}$	17 $\mu\text{l}$
	Eau ultra-pure	13 $\mu\text{l}$	221 $\mu\text{l}$
	Volume Total	19 $\mu\text{l}$	323 $\mu\text{l}$

Dans cette étape-là, un témoin positif est ajouté, il nous renseigne sur le bon fonctionnement du kit « Ready Reaction Mix », en effet dans la boîte de ce réactif en plus de trouver un flacon du « Ready Reaction Mix », deux autres flacons sont présents, un contenant l'ADN contrôle « pGEM®-3Zf (+) double-stranded DNA Control Template » et son primer « -21 M13 Control

Primer » qui est un primer Forward. Dans un micro-tube de 0.2ml noté « T+RM » nous avons déposé : 10.5 µl d'eau ultra pure, 4 µl du Sequencing Buffer, 4 µl du primer de l'ADN contrôle, et 0.5 µl de l'ADN contrôle.

Les tubes contenant le volume réactionnel sont ensuite déposés sur la plaque du thermocycleur et le programme sequencing est lancé comme indiqué dans le **tableau VI**.

**Tableau VI.** Programme du Thermocycleur pour le séquençage.

Etape de Dénaturation Initiale	25 Cycles			Etape de Maintenance
	Dénaturation	Hybridation	Extension (Elongation)	
Maintien	Cycle			Maintien
96°C, 1 minute	96°C, 10 secondes	50°C, 5 secondes	60°C, 4 minutes	4°C, ∞

## II.9. Purification Post-Séquençage

### II.9.1. Principe

Suite au séquençage, certains ddNTPs marqués ne seront pas incorporés ainsi que certains sels libérés au cours des cycles de séquençage doivent être éliminés afin de purifier l'échantillon d'ADN. Ceci dans le but d'éviter et de limiter au maximum la co-injection de ses impuretés avec l'ADN à analyser lors de l'électrophorèse capillaire. Différentes techniques de purification existent dont la purification par précipitation à l'éthanol et la purification sur colonnes. Par contre la technique la plus avantageuse du point de vue simplicité, rapidité et le coût qui est moindre reste la purification à l'aide du kit Big-Dye X-Terminator. Cette purification se base sur l'utilisation de seulement deux solutions ayant des propriétés bien définies.

- ❖ La solution X-Terminator a la faculté de capturer les ddNTPs marqués non incorporés et autres particules.
- ❖ La solution SAM permet d'améliorer l'action de la solution X-Terminator et de stabiliser l'échantillon après purification.

Ainsi, pour une bonne purification, l'agitation permet à l'action de la solution X-Terminator d'être à son maximum, alors que lors de la centrifugation, cette même solution ainsi que les composés capturés sont déplacés vers le fond du puit (culots) et le produit purifié, se retrouve déplacé en haut du puit (surnagent) ([PROTOCOL SUMMARY BigDye® XTerminator™ Purification Kit, 2006](#) ; [Applied Biosystems BigDye® XTerminator™ Purification Kit, Protocol, 2007](#)).

### II.9.2. Mode opératoire

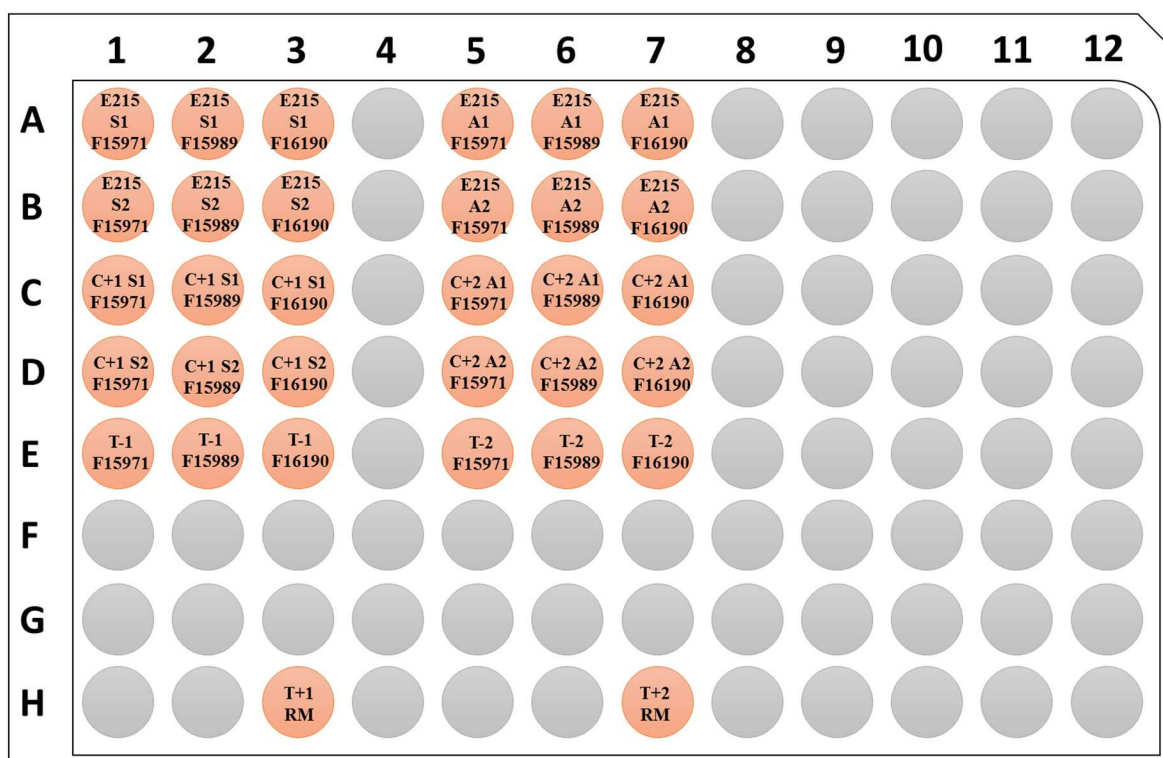
Après la réaction de séquençage, l'étape de purification des produits séquencés à l'aide du kit « Big-Dye X-Terminator » est la suivante. Commencer par mélanger 45 µl de la solution

SAM et 10  $\mu\text{l}$  de la solution X-Terminator représentant la solution de purification (**Tableau VII**).

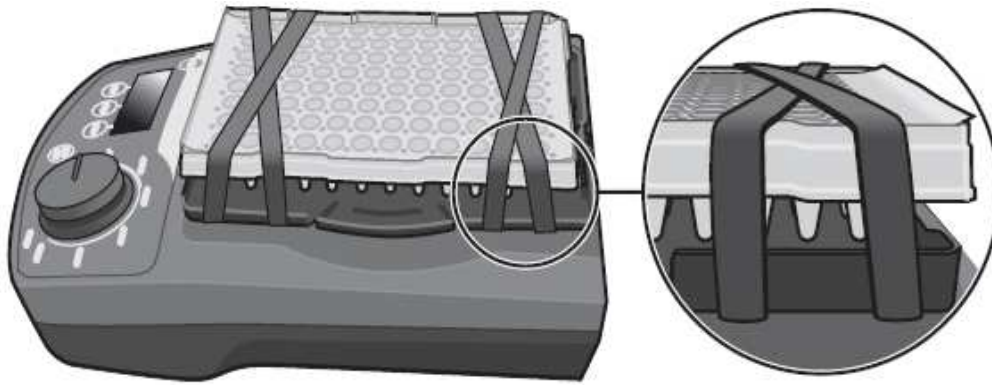
**Tableau VII.** Calculs des volumes du mélange réactionnel de la solution de purification

		Quantité par nombre d'échantillons	
		1 échantillon	17 échantillons + 1
Composants du Mélange Réactionnel	Solution SAM	45 $\mu\text{l}$	765 $\mu\text{l}$
	Solution X-Terminator	10 $\mu\text{l}$	170 $\mu\text{l}$
	Volume Total	55 $\mu\text{l}$	935 $\mu\text{l}$

55  $\mu\text{l}$  du mélange vortexé, sont déposés dans des puits d'une plaque de 96 puits, auxquels sont ajoutés 10  $\mu\text{l}$  d'ADNmt séquençé. Les puits vides sont remplis de 10 $\mu\text{l}$  d'eau-ultra pure. Ensuite la plaque est scellée avec un film transparent et mise sur l'agitateur adapté en la fixant avec des élastiques. L'agitation se fait avec le mode B à 2000 rpm pendant 30 minutes. Une fois l'agitation terminée, centrifuger la plaque à 1000g pendant 2 minutes, ensuite retirer délicatement le film transparent et sceller, de nouveau, la plaque avec un Septa stérilisé.



**Figure 13.** Exemple d'un plan de plaque utilisé



**Figure 14.** Placement de la plaque sur l'agitateur (Applied Biosystems BigDye® XTerminator™ Purification Kit, Protocol, 2007).

## II.10. Electrophorèse Capillaire

### II.10.1. Principe

L'électrophorèse capillaire est une technique d'analyse permettant la séparation des fragments d'ADN ou autres particules en se basant sur la différence de leur poids moléculaire, ceci en présence d'un champ électrique. La migration se fait tout au long des capillaires tapissés de polymère, de l'anode vers la cathode. Cet appareil est aussi muni d'une fenêtre de détection qui, au bout de la migration des fragments et à l'aide d'un laser excitera les fragments marqués qui émettront une fluorescence qui sera capté par la caméra jointe à la fenêtre de détection. Ainsi les résultats seront reportés et visualisés sur un chromatographe (Butler, 2005 ; Butler, 2011 ; Shewale et Liu, 2016).

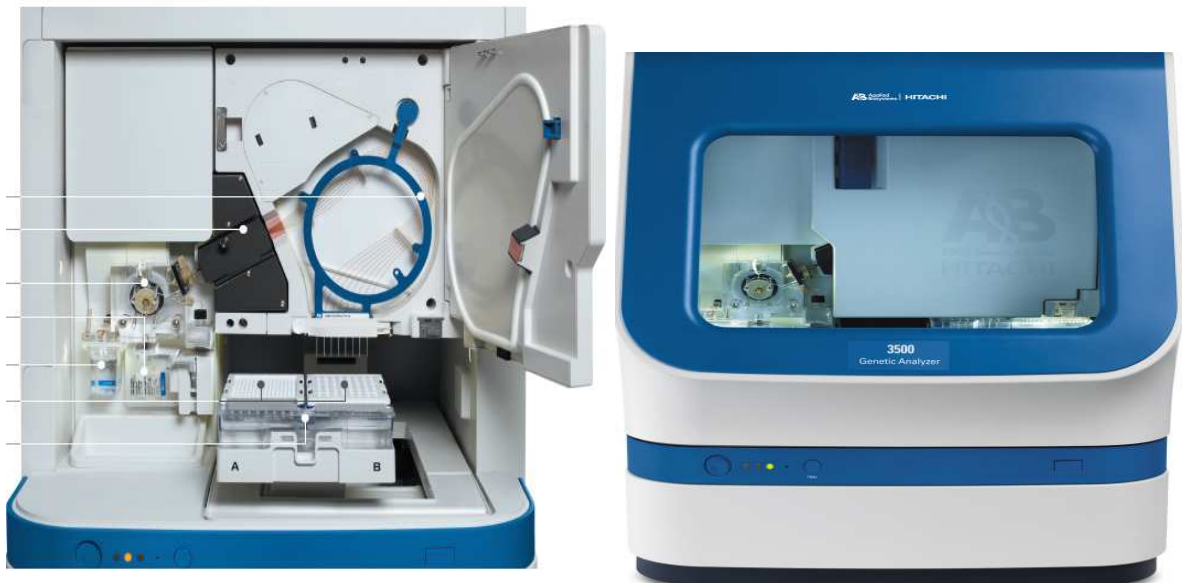
### II.10.2. Mode opératoire

L'électrophorèse capillaire a été réalisée par le Séquenceur « 3500xl Genetic Analyzer d'Applied Biosystems ». Le logiciel « DATA Collection » de ce dernier, a été utilisé pour saisir le nom de la plaque, le type d'analyse et de vérifier ses paramètres adéquats (**Tableau VIII**). Ainsi qu'à éditer le plan de plaque et de sélectionner trois systèmes afin d'attribuer et d'organiser le report des résultats

La plaque de 96 puits contenant les produits séquencés, purifiés a été scellée d'un Septa et placée sur le support à plaque du séquenceur 3500xl, puis déposée au niveau du Séquenceur 3500xl sur le champ destiné « A » ou « B » (Autosampler). Enfin la plaque a été attribuée à sa position, l'analyse a été lancée et les résultats ont été récupérés après un certain temps.

**Tableau VIII.** Paramètres du Séquenceur 3500xl spécifique au Protocole étudié

Paramètres du Séquenceur 3500xl	Attribution spécifique
Nom du protocole :	BDX Rapid seq 36_POP4xl 1
Type d'analyse :	Sequencing
Taille des Capillaire :	36cm
Polymère :	POP4
Type de plaque :	96 puits
Température du Four :	55C°
Voltage :	13.6 kV
Voltage pré Run :	15 kV
Voltage d'Injection :	1.2 kV
Temps du Run :	1700sec
Temps pré Run :	180sec
Temps d'Injection :	18sec
Délai de transmission des données (Data Delay) :	240sec

**Figure 15.** Séquenceur 3500xl Genetic Analyzer d'Applied Biosystems.

## II.11. Lecture des Données

### II.11.1. Sequencing Analysis Software v6.0 Applied Biosystems™

#### II.11.1.1. Principe

Le Sequencing Analysis est un logiciel d'analyse doté d'algorithmes permettant d'analyser, d'afficher, d'éditer et de stocker (format électronique ou papier) tous les résultats et données brutes obtenus par l'électrophorèse capillaire des séquenceurs d'ADN (ex : Séquenceur 3500xl Genetic Analyzer d'Applied Biosystems). Ainsi ce logiciel permet de faciliter la visualisation et l'interprétation des données et donc d'appuyer l'analyse d'identification génétique ([User Bulletin, Sequencing Analysis Software v5.3, 2007](#) ; [Applied Biosystems Sequencing Analysis Software Quick Reference Card, 2009](#)).

#### II.11.1.2. Mode opératoire

La lecture des résultats du séquençage des échantillons analysés a été effectuée par le logiciel « Sequencing Analysis Software v6.0 Applied Biosystems™ ». La visualisation des résultats a commencé un à un (par primer) et elle s'est déroulée comme suit :

- ❖ Le « Raw Data » : les données brutes du séquençage de l'échantillon,
- ❖ L'électrophorégramme : représentant les pics de chaque base de la séquence de l'échantillon,
- ❖ L'annotation : qui illustre l'intensité de chaque base.

De cette façon l'analyse a été répétée pour chaque primer de chaque échantillon analysé.

### II.11.2. Sequencher® v5.4.2

#### II.11.2.1. Principe

Le Sequencher® est un logiciel de bio-informatique conçu par la société « Gene Codes Corporation », utilisé pour analyser et assembler des séquences d'ADN. Ces dernières sont de courtes longueurs et ont la particularité d'être voisines ou proches l'une à l'autre (Contigüe). De ce fait le logiciel « Sequencher® » permet de générer une séquence d'ADN continue et ordonnée suite à un assemblage. Cette séquence obtenue porte le nom de « Contig ». Le Sequencher® permet aussi de comparer le Contig à des séquences référence, de visualiser les différences et ainsi d'obtenir un profil type de chaque échantillon analysé ([Sequencher® 5 Series, User Manual, Gene Codes Corporation, 2015](#)).

#### II.11.2.2 Mode opératoire

Pour la dernière étape du protocole, une autre lecture des données représentant les séquences lues par chaque primers pour chaque échantillon, a été réalisée par le logiciel « Sequencher® 5.4.2 ». Pour chaque échantillon, la visualisation des données a été faite par primer (un à un), où une 1<sup>ère</sup> vérification du début et de la fin du chromatogramme a été effectuée, par l'édition des pics non représentatifs.

Suite à cela, un assemblage des séquences de chaque primer, en se référant à la « rCRS » a été appliqué, donnant une séquence de la « CR » qui est continue et ordonnée. Cette dernière a

pu être comparée à la « rCRS » en visualisant chaque pic de chaque base de la séquence analysée, afin d'identifier les points de différence.

## **II.12. Paramètre de validation**

Afin d'accepter les trois critères de validation de notre étude qui sont : la reproductibilité, la répétabilité et l'exactitude, les procédures suivantes ont été effectuées. Chaque étape du protocole d'analyse a été réalisée par deux manipulateurs différents « S » et « A » à des temps différents mais sur la même série des seize échantillons (Reproductibilité). Chaque manipulateur a effectué l'analyse des 16 échantillons en deux fois et en simultané « S1-S2 » et « A1-A2 » (Répétabilité). Enfin, les échantillons à analyser contenaient trois ADN contrôle (T+, C+, M+) dont le profil mitochondrial de la « CR » est connu, suite à leur analyse, les résultats ont été comparés aux résultats de références (Exactitude).



# **Chapitre III :**

## **Résultats et**

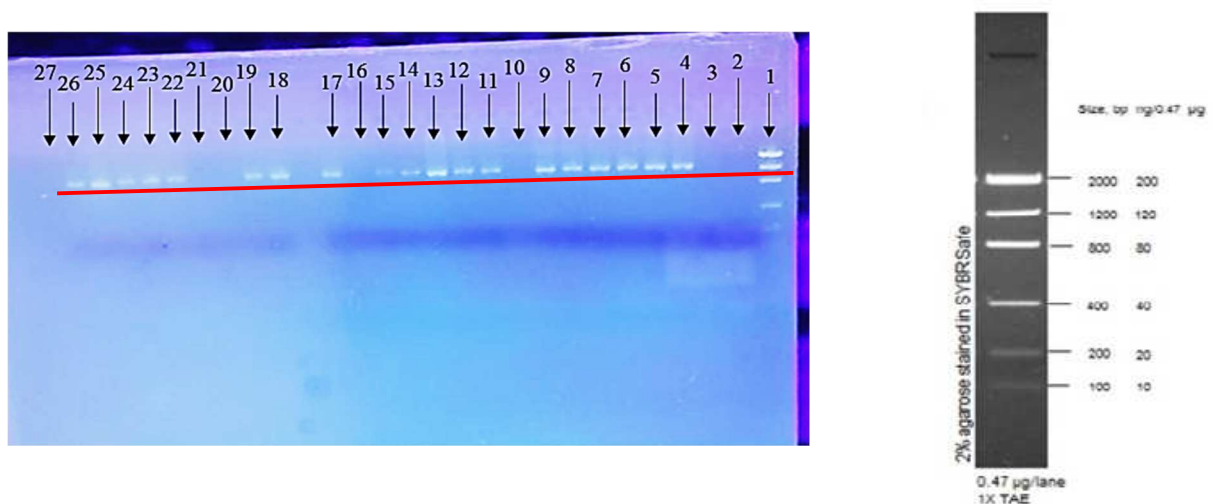
### **discussion**

## Chapitre III : Résultats et discussion

### III.1. Résultats de la quantification

Les produits amplifiés lors de cette étude, ont été analysés par électrophorèse sur gel d'agarose pour une quantification de l'ADNmt. Les résultats sont visualisés par un transilluminateur UV, qui apparaissent sous forme de bandes fluorescentes. Ces résultats nous renseignent à la fois sur la taille des fragments d'ADNmt amplifié, en observant la distance de migration, et sur la quantité approximative de l'ADNmt en se servant de l'intensité de fluorescence.

En comparant les bandes obtenues pour chaque échantillon avec l'échelle « Low Mass DNA Ladder », il a été déduit que la taille des fragments était de 1200 pb pour tous les échantillons, ce qui correspondait au résultat attendu puisque la taille de la CR est de 1200 pb. L'intensité de chaque bande a indiqué que la quantité de l'ADNmt est de 120 ng approximativement, comme illustré dans **la figure 16**.



**Figure 16.** Exemple d'électrophorèse sur gel d'agarose

**Puit 1** : Ladder, **Puit 2** : E9S, **Puit 3** : E17S, **Puit 4** : E18S, **Puit 5** : E115S, **Puit 6** : E151S, **Puit 7** : E172S, **Puit 8** : E195S, **Puit 9** : E231S, **Puit 10** : E234S, **Puit 11** : E630, **Puit 12** : E631S, **Puit 13** : M+1, **Puit 14** : T-1, **Puit 15** : E9A, **Puit 16** : E17A, **Puit 17** : E18A, **Puit 18** : E115A, **Puit 19** : E151A, **Puit 20** : E172A, **Puit 21** : E195A, **Puit 22** : E231A, **Puit 23** : E234A, **Puit 24** : E630A, **Puit 25** : E631A, **Puit 26** : M+2, **Puit 27** : T-2

Tableau IX. Résultats de la quantification des produits amplifiés

Echantillon	Quantité de l'ADN (ng)		Taille des fragments (pb)	
	Manipulateur S	Manipulateur A	Manipulateur S	Manipulateur A
<b>E9</b>	Indéetectable	120	Indéetectable	1200
<b>E10</b>	120	120	1200	1200
<b>E17</b>	Indéetectable	Indéetectable	Indéetectable	Indéetectable
<b>E18</b>	120	120	1200	1200
<b>E115</b>	120	120	1200	1200
<b>E151</b>	120	120	1200	1200
<b>E172</b>	120	Indéetectable	1200	Indéetectable
<b>E195</b>	120	Indéetectable	1200	Indéetectable
<b>E215</b>	120	120	1200	1200
<b>E231</b>	120	120	1200	1200
<b>E234</b>	Indéetectable	120	Indéetectable	1200
<b>E630</b>	120	120	1200	1200
<b>E631</b>	120	120	1200	1200
<b>T+</b>	120	120	1200	1200
<b>C+</b>	120	120	1200	1200
<b>M+</b>	120	120	1200	1200
<b>T-</b>	120	Indéetectable	1200	Indéetectable

Le **tableau IX** comprend les résultats de la quantification. Sur les 17 échantillons quantifiés nous avons obtenus :

- Une visualisation de 11 échantillons confirmant une taille de 1200 pb et une quantité de 120 ng. Ce résultat nous permet d'attester du bon déroulement des deux étapes d'extraction et d'amplification.
- Un problème de détection a été enregistré pour 4 échantillons, qui étaient détectés soit par le manipulateur « A » soit par le manipulateur « S » mais pas les deux en même temps. Un des échantillons n'a été détecté ni pour le manipulateur « A » ni pour le manipulateur « S ». Ces problèmes de détection seraient probablement dus au fait que certaines cartes FTA utilisées avaient un fond blanc sans indicateur coloré d'application de l'échantillon où les poinçonnages ont pu être opérés sur des zones ne contenant pas suffisamment d'ADN.

En outre, le T<sup>-</sup> du manipulateur « A » a été indéetectable ce qui confirme qu'il n'y a pas eu de contamination lors de la manipulation. Alors que, la bande du T<sup>-</sup> du manipulateur « S » est apparu en raison d'une erreur qui s'est produite au cours de la mise en tube des poinçons où le poinçonnage a été réalisé sur une carte FTA contenant un échantillon au lieu d'une carte FTA vide. De ce fait, les échantillons qui n'ont pas été détectés ont été éliminés de l'analyse des étapes ultérieures.

L'absence d'étape de lavage des poinçons lors de l'extraction a conduit à une non détection des échantillons lors de la quantification. Ce qui nous a amené à confirmer nos essais en procédant à des extractions en présence et en absence de lavage. La comparaison des deux variantes est illustrée dans le **tableau X**.

**Tableau X.** Résultats de la quantification des échantillons sans lavage et avec lavage

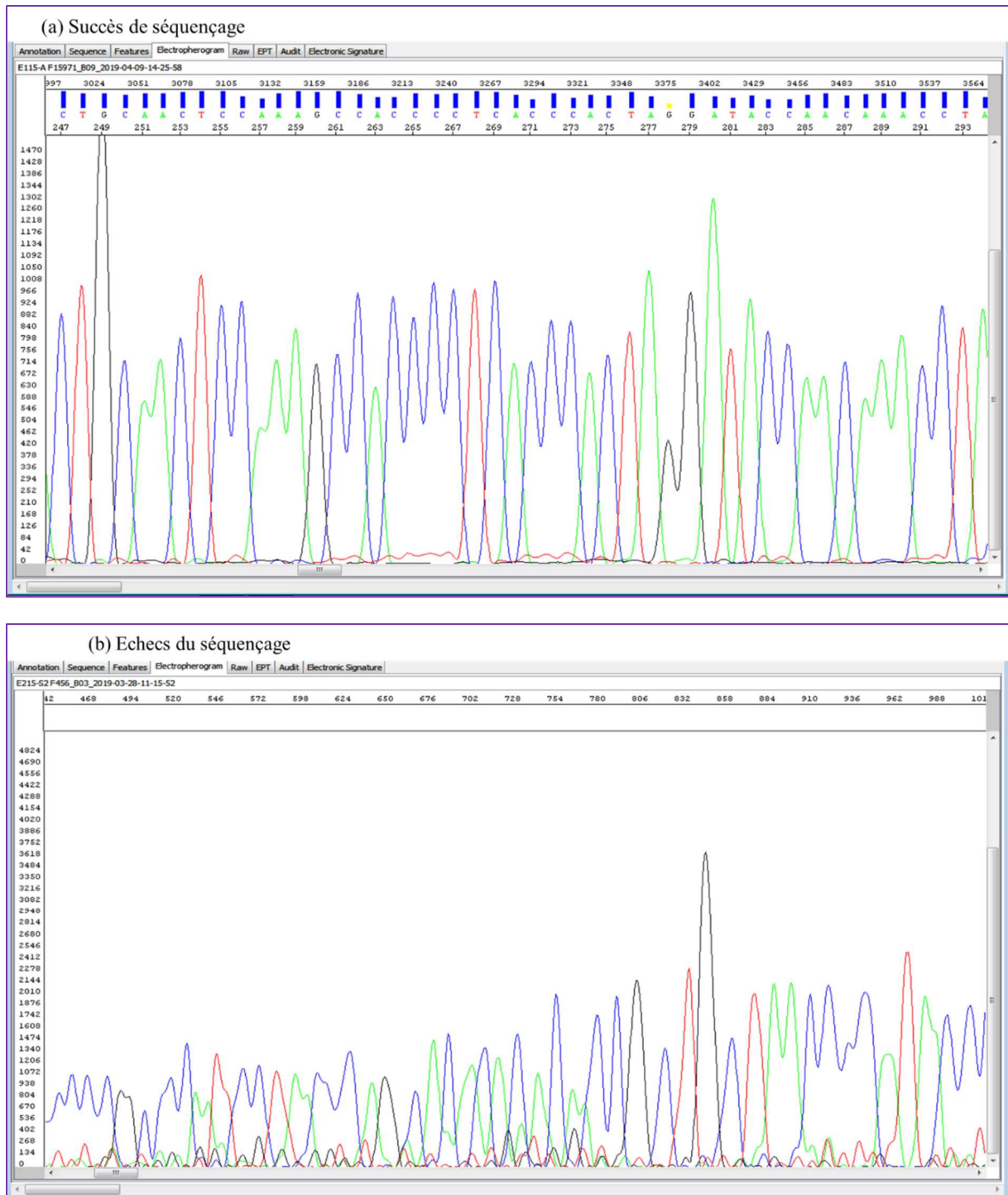
	Manipulateur S		Manipulateur A	
	Sans lavage	Avec lavage	Sans lavage	Avec lavage
<b>E10</b>	Indétectable	Détectable	Indétectable	Détectable
<b>E215</b>	Indétectable	Détectable	Indétectable	Détectable

## III.2. Résultats du séquençage

### III.2.1. Sequencing Analysis Software v6.0 « SeqA »

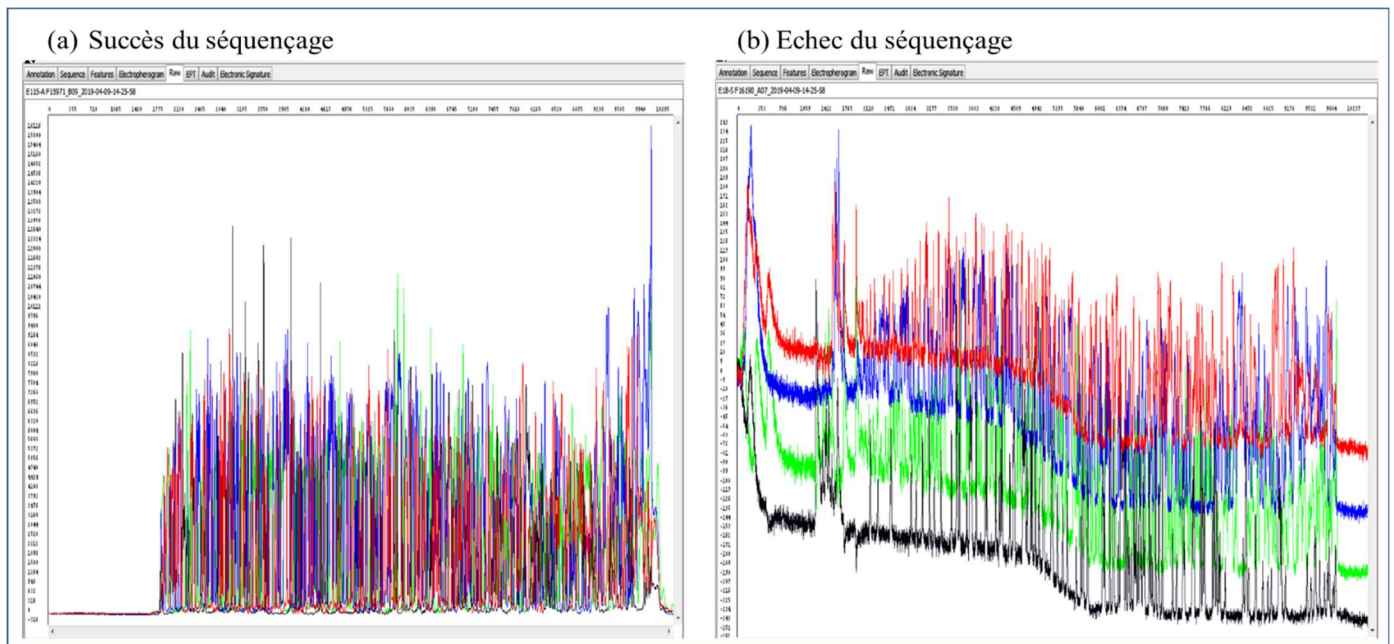
Après la réaction de séquençage et migration des séquences par l'électrophorèse capillaire, la vérification du succès de cette réaction et la visualisation des résultats de chaque amorce pour chaque échantillon ont été réalisées par le logiciel « SeqA Software 6 ». Ce dernier, permet de lire les données sous différents aspects :

❖ L'électrophorégramme nous a présenté l'enchaînement des bases azotées ainsi que leur position dans le brin sens et antisens de la séquence lue, sous forme de pics de quatre couleurs différentes (A = vert, G = noir, T = rouge et C = bleu) et d'intensité variable. La couverture de lecture des primers a également été déterminée à ce niveau. **La figure 17** montre deux exemples d'électrophorégramme obtenus après succès **(a)** et échecs **(b)** de la réaction de séquençage.



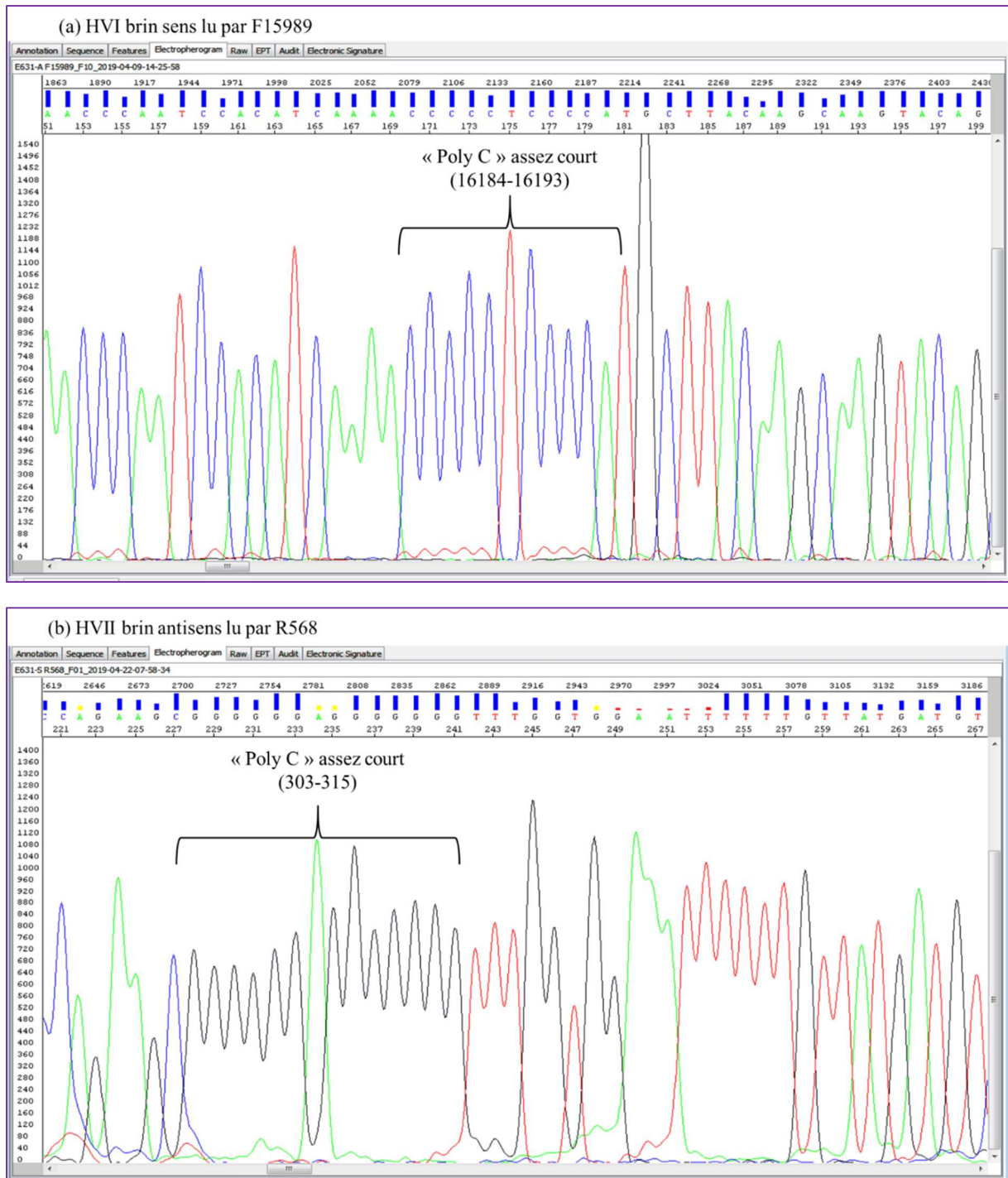
**Figure 17.** Exemples de résultats d'électrophorégramme obtenus après une lecture des séquences analysées

❖ Le « Raw Data » représentant un spectre des données brutes du résultat indiquant la qualité de déroulement de la réaction de séquençage en visualisant le signal de fluorescence pour chaque base azotée (A, G, T et C) ainsi que l'intensité de ce signal qui montre le seuil maximal illustrant un aperçu sur la quantité d'ADN utilisé dans la solution.

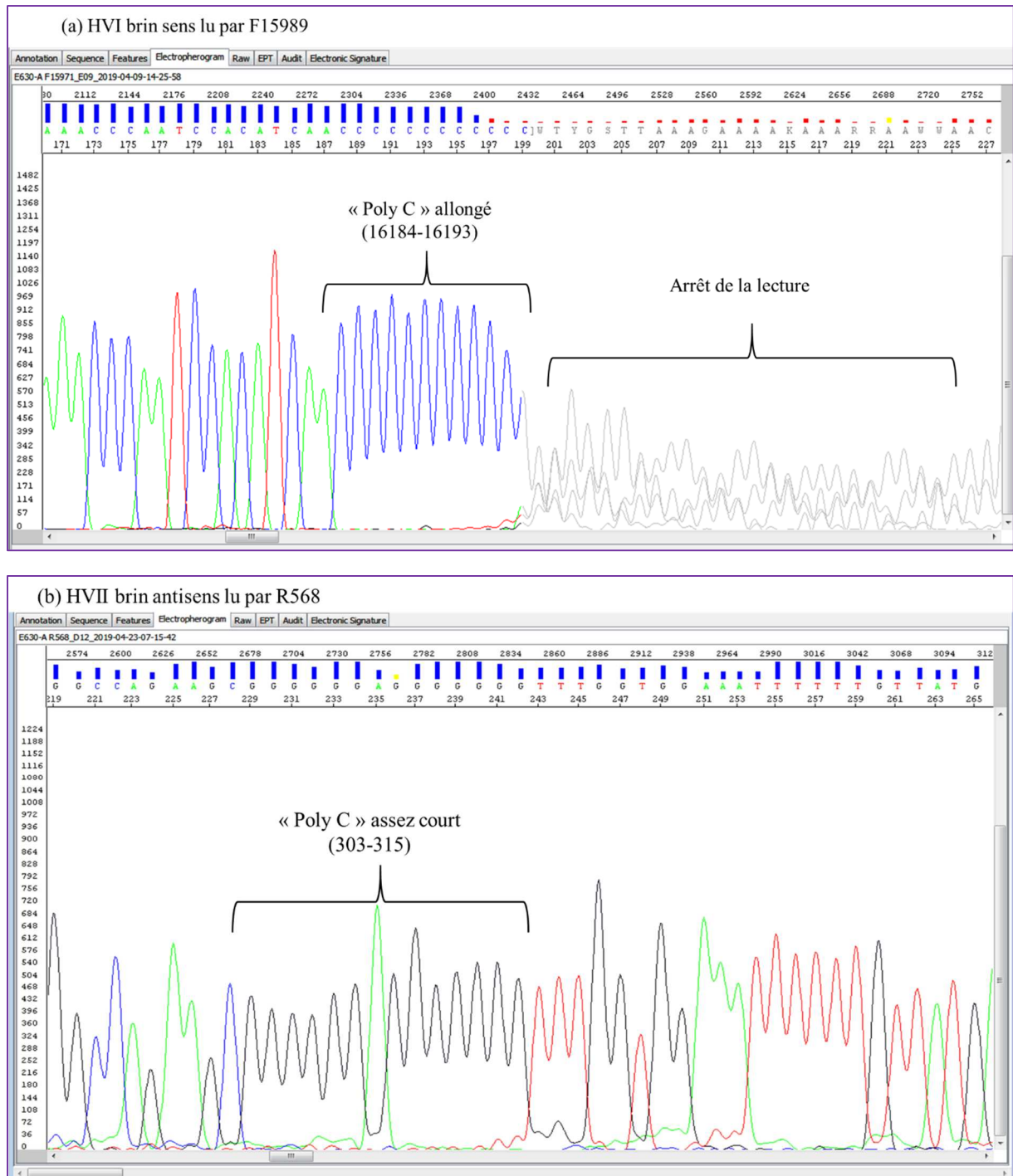


**Figure 18.** Exemple de la lecture des résultats sous forme de « Raw Data »

Après analyse des échantillons, il a été constaté qu'au niveau d'un « Poly C » assez court dans les régions HVI et/ou HVII, la lecture par les amorces ne s'arrête pas et continue jusqu'à arriver à sa capacité maximale (**Figure 19**), alors qu'au niveau d'un « Poly C » allongé dans les régions HVI et/ou HVII la lecture par les primers s'arrêtent à ce niveau et se poursuit par le primer se situant après cette région (**Figure 20** et **Figure 21**).

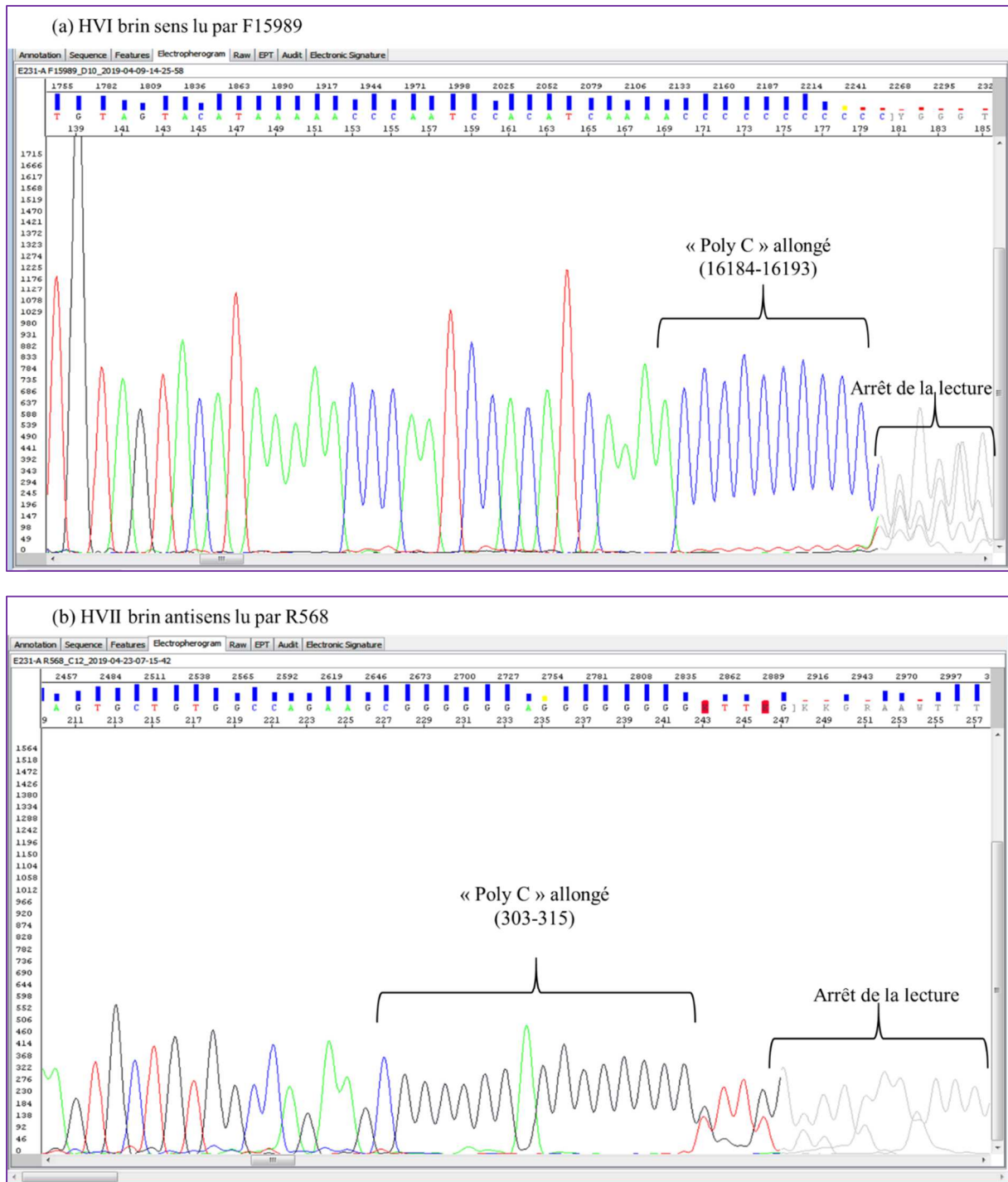


**Figure 19.** Exemple de résultat de séquençage de HVI du brin sens et HVII du brin antisens de l'échantillon E631 contenant un « Poly C » court dans les deux régions.



**Figure 20.** Exemple de résultat de séquençage de HVI du brin sens et HVII du brin antisens de l'échantillon E630 contenant un « Poly C » court et un allongé

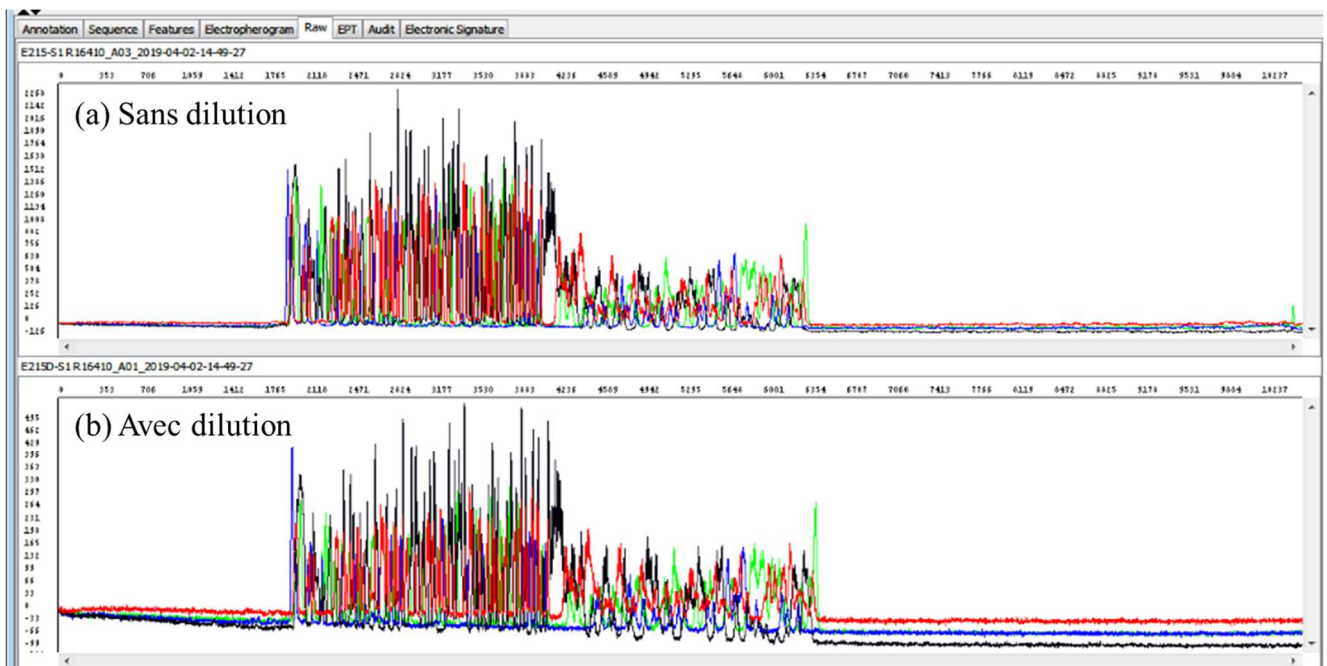




**Figure 21.** Exemple de résultat de séquençage de HVI du brin sens et HVII du brin antisens de l'échantillon E231 contenant un « Poly C » allongé dans les régions.

❖ **Pré-séquénçage :** En se basant sur ces résultats, nous avons proposé une étape de pré-séquénçage où l'échantillon est testé par les amorces F1591 ou F15989 et R614 ou R568 pour vérifier le type de « Poly C » qui se trouve dans les régions HVI et HVII respectivement afin de choisir le design des primers qui convient.

Après purification et avant séquençage, le protocole préconise une étape de dilution (Cf. Chap II, p 22). Au cours des essais nous avons comparé les résultats de séquençage en présence ou non de l'étape de dilution (**Figure 22**)



**Figure 22.** Raw Data de l'échantillon E215 dilué et non dilué séquençé avec l'amorce R16410

Les résultats obtenus de deux séries d'essais ont montré que le séquençage de l'échantillon non dilué était de meilleure qualité que celui de l'échantillon dilué. L'observation du spectre du « Raw Data » a indiqué que pour l'analyse avec dilution, le signal de fluorescence de chaque base s'est montré de qualité moindre par rapport à celui l'analyse sans dilution. De même, l'intensité du signal a été très faible pour l'échantillon dilué et parfois peut arriver à des valeurs négatives alors que pour l'échantillon non dilué l'intensité a été de valeurs élevées montrant ainsi des résultats favorables à l'analyse. Suite à ce bilan, les échantillons ont été étudiés sans passer par l'étape de dilution pour la validation du protocole.

Lors de l'analyse des résultats de l'électrophorogramme des témoins négatifs T<sup>-</sup>, certains pics ont été observables mais non interprétables car au niveau de la séquence aucun étiquetage n'a été observé (aucune base nucléique n'a été attribuée) (**Figure 23**). Ces pics, représentent la détection du bruit de fond issu de l'électrophorèse capillaire en raison de la précision de lecture du séquenceur utilisé (3500xl).

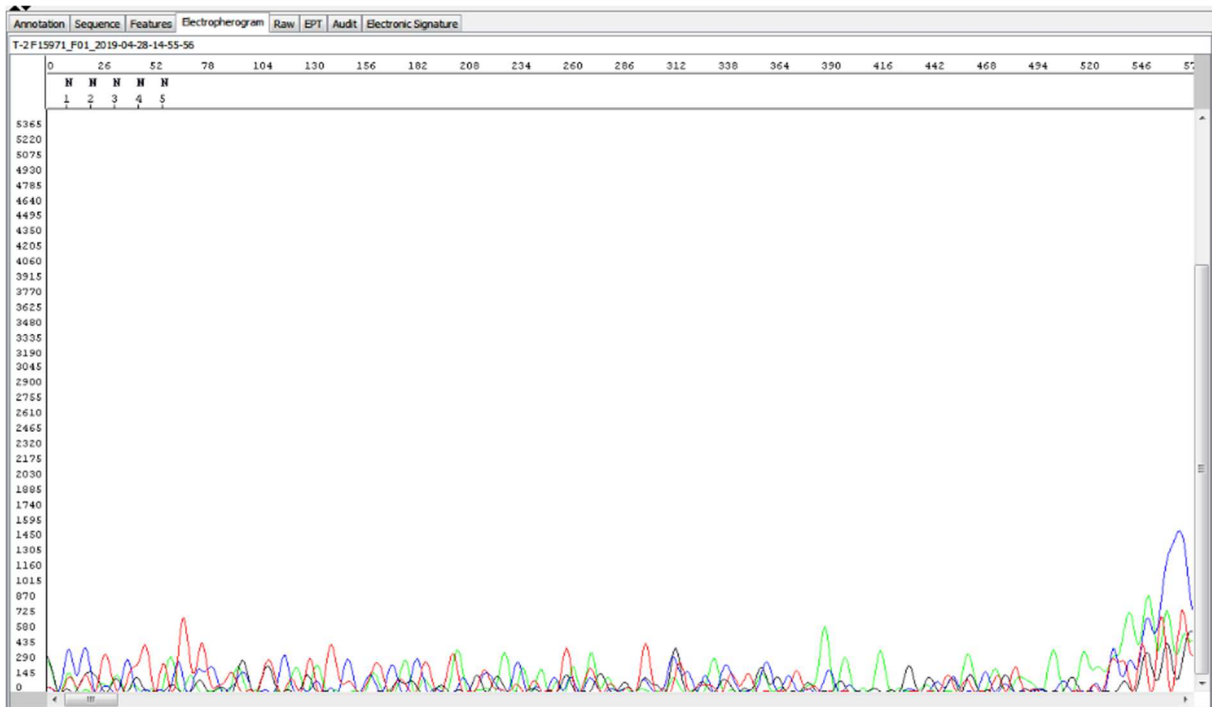


Figure 23. Electrophorégramme du T<sup>r</sup> F15971

Concernant le Raw Data d'un T<sup>r</sup>, il a été illustré de façon différente à celui d'un ADN analysé convenablement. Il a été remarqué que le signal de fluorescence pour chaque base ne constitue pas une ligne droite horizontale, chaque signal tend vers le bas séparément des autres (Figure 24).

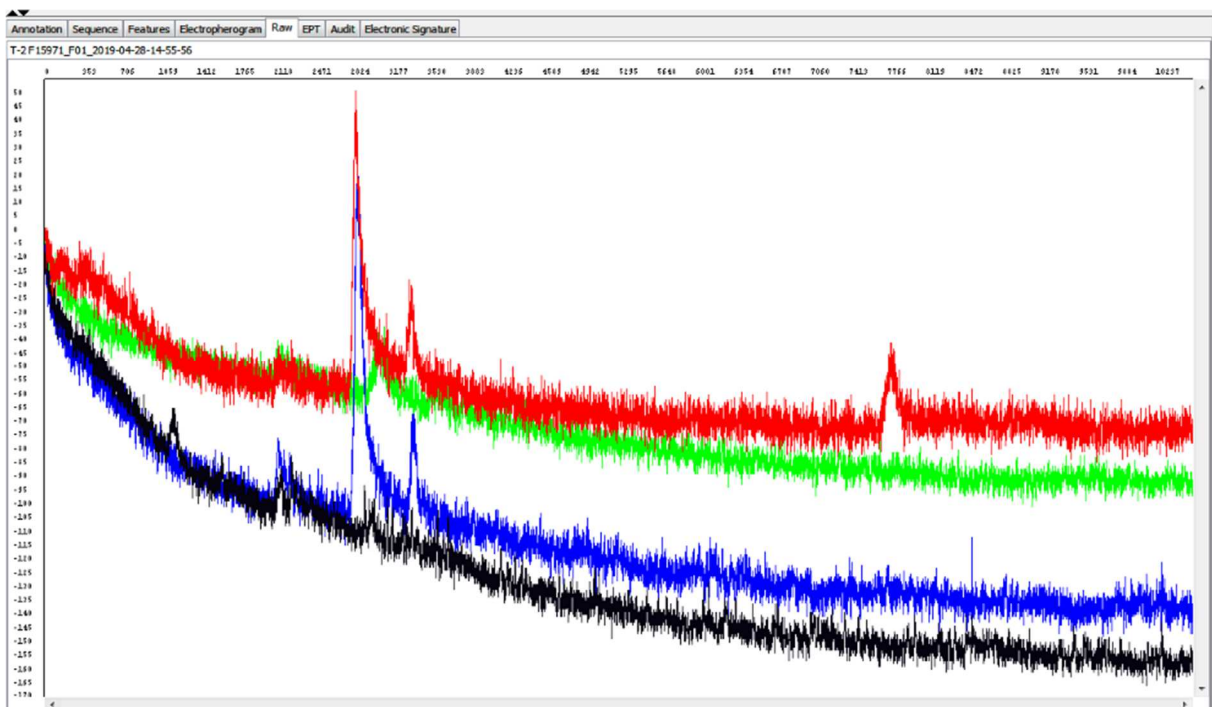


Figure 24. Données brutes « Raw Data » du T-2 F15971

Suite à la lecture des résultats par le logiciel « Sequencing Analysis Software v6.0 », tous les échantillons ont été séquencés avec succès et nous avons pu déterminer le type de Poly C trouvé dans les régions hypervariables (HVI et HVII) pour chaque échantillon (**Tableau XI**), sauf pour l'échantillon E151 où après plusieurs essais par différents manipulateurs, ce dernier n'a pas abouti à des résultats de séquençage convenables. Il a été conclu que cet échantillon était défaillant donc il a été éliminé de l'analyse.

**Tableau XI.** Résultats de séquençage lu par « SeqA »

	HVI		HVII	
	Poly C court	Poly C allongé	Poly C court	Poly C allongé
<b>E9</b>	+	-	+	-
<b>E10</b>	+	-	+	-
<b>E18</b>	+	-	+	-
<b>E115</b>	+	-	+	-
<b>E172</b>	+	-	+	-
<b>E195</b>	-	+	+	-
<b>E215</b>	-	+	+	-
<b>E231</b>	-	+	-	+
<b>E234</b>	+	-	+	-
<b>E630</b>	-	+	+	-
<b>E631</b>	+	-	+	-
<b>T+</b>	+	-	+	-
<b>C+</b>	-	+	+	-
<b>M+</b>	+	-	+	-

+ : présence

- : absence

Les résultats de SeqA nous renseignent sur différents paramètres importants pour l'étape suivante de l'analyse à savoir : la nature de l'échantillon analysé, le fonctionnement du réactif du séquençage, les bruits de fond (dus à des erreurs de purification ou à des problèmes électrophorétiques) ainsi que l'intensité des pics.

### III.2.2. Résultats du Sequencher

Suite à la visualisation des données du séquençage pour chaque échantillon par le logiciel précédent, une seconde lecture a été réalisée par le Sequencher. En suivant le mode opératoire déjà cité, le contig de la séquence de la « CR » est obtenu. Les différents résultats possibles de ce logiciel sont : l'aperçu du contig, le chromatogramme et le rapport du profil ADNmt. Ils sont présentés de la sorte :

1. Aperçu du contig « Summary » : au niveau de ce résultat une vue globale de la séquence entière, assemblée et continue de la région de contrôle est présentée avec un code couleur représentant le type de lecture (double, sens inverse etc.). Ainsi que les différents primers qui ont été utilisés lors du séquençage pour l'obtention de la séquence,

positionnés à l'endroit même de leur hybridation. Une distance plus ou moins approximative du nombre de base lu par chaque primer est aussi mentionnée.

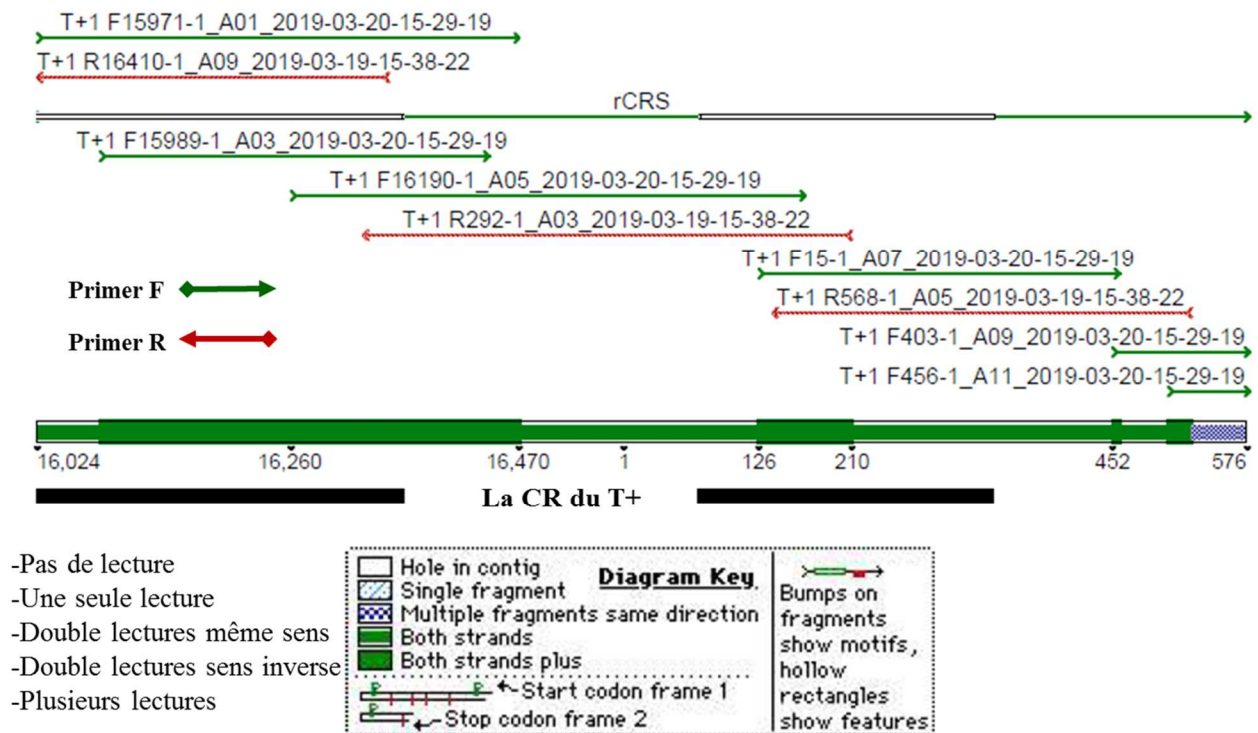
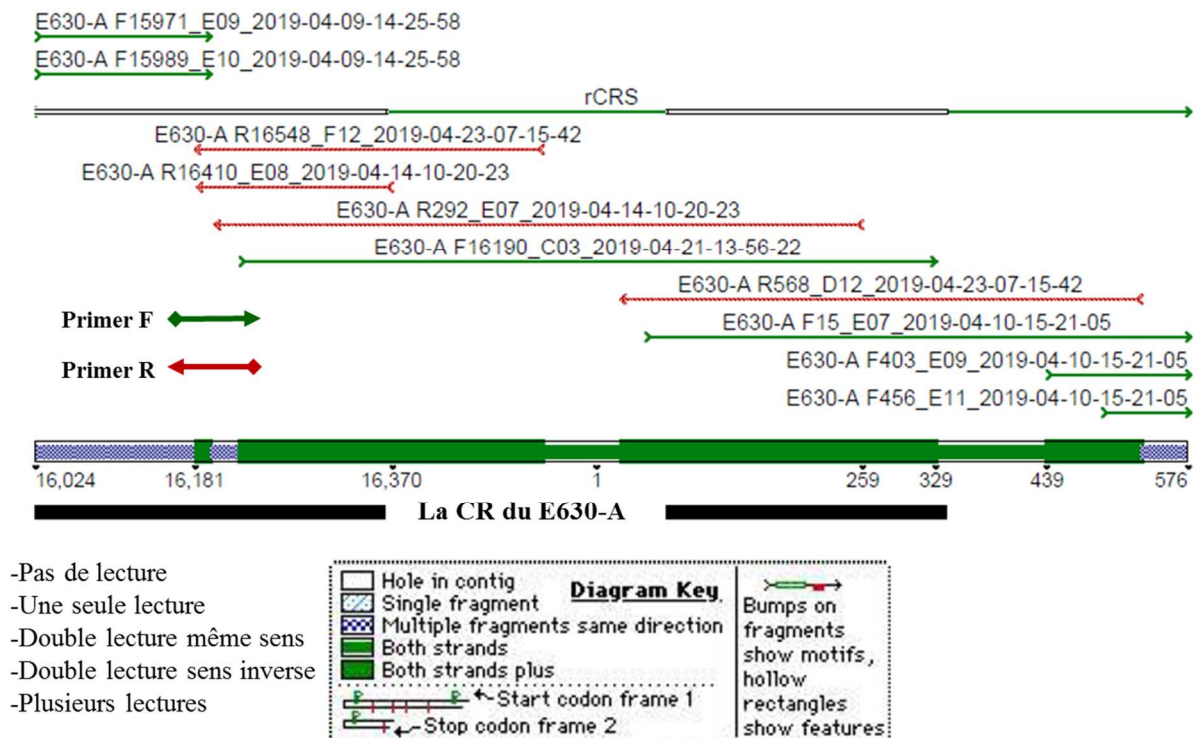


Figure 25. La CR du T+ assemblée.

La **figure 25** illustre la région de contrôle séquencée de l'ADNmt standard T+ suite à l'assemblage des séquences obtenues à partir du séquençage effectué avec le 2<sup>ème</sup> design (9 primers). Le choix de ce dernier a été possible en se référant aux résultats du standard de référence « SRM 2392 » concernant la nature de HVI et HVII de l'ADNmt « 9947A ». Ainsi, les 9 primers ont permis de couvrir la totalité de la CR de la meilleure façon possible. Selon la figure :

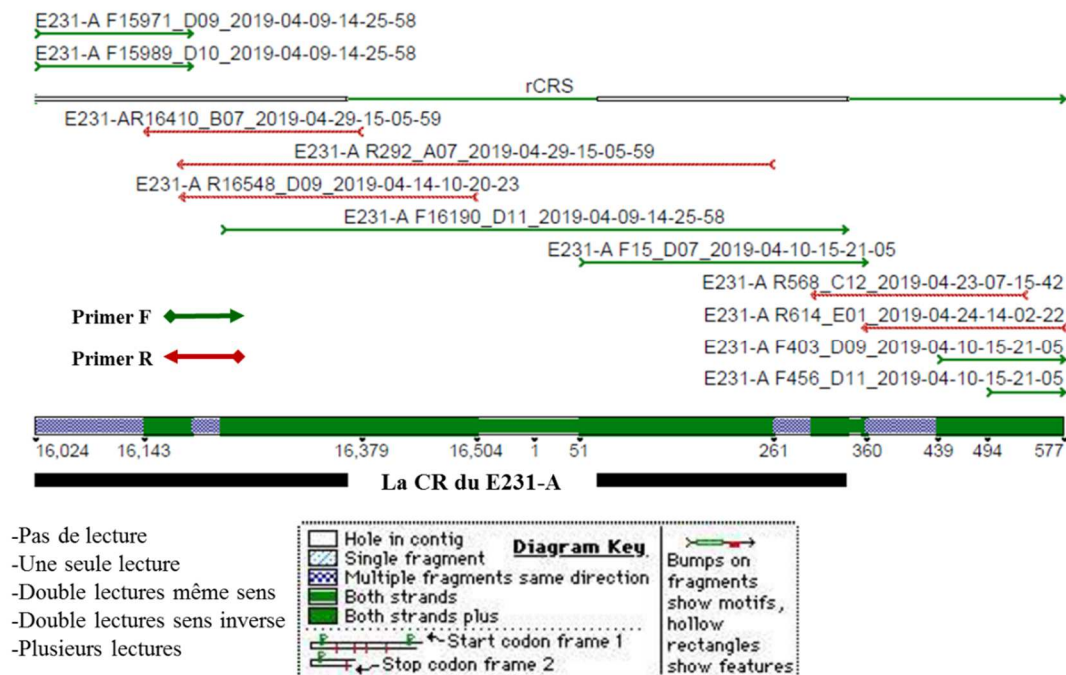
- De « 16024 » à « 520 » : il y'a eu une double ou plusieurs lectures dans les deux sens (F et R),
- De « 520 » à « 576 » : il y'a eu une double lecture avec les primers F (même sens).



**Figure 26.** La CR du E630-A assemblée

Le résultat d'un assemblage des fragments séquencés de la région de contrôle en utilisant le 3<sup>ème</sup> design (10 primers), est illustré dans la **figure 26**. Le E630 a d'abord été séquencé en utilisant le 2<sup>ème</sup> design, mais en raison de la présence d'un « Poly C » allongé au niveau de la HVI, les primers F15971 et F15989 s'arrêtent brusquement. De ce fait, le primer R16548 a été rajouté (3<sup>ème</sup> design) ce qui a permis de combler la lecture et de couvrir la « CR » en sa totalité. Par rapport à cet échantillon la lecture est de types :

- Double lecture dans le même sens (primer F) : de « 16024 » à « 16181 », de « 16190 » à « 16210 » et à la fin de la séquence,
- Double ou plusieurs lectures dans les sens inverse (F et R) : de « 16181 » à « 16190 », de « 16210 » à « 560 » environ.



**Figure 27.** La CR du E231-A.

La **figure 27** a été obtenue suite à plusieurs tentatives de séquençage afin d'avoir un assemblage des fragments ayant été couverts en totalité avec des doubles lectures au minimum. En utilisant le 2<sup>ème</sup> design, les régions hypervariables n'ont pas pu donner l'assemblage souhaité et les primers s'arrêtaient de façon subite en raison des « Poly C » allongés. De ce fait en plus d'avoir rajouté le R16548 (3<sup>ème</sup> design), nous avons rajouté un autre primer Reverse le R614 (4<sup>ème</sup> design), ce qui nous a permis d'avoir un assemblage représentatif du contig du E231 essentiellement au niveau de ses régions hypervariables où le souci se posait. La lecture du E231 s'est faite comme suit :

- De « 16024 » à « 16143 », de « 16250 » à « 16260 », de « 261 » à « 280 », et de « 360 » à « 439 » : une double lecture dans le même sens,
- De « 16143 » à « 16250 », de « 16260 » à « 261 », de « 280 » à « 360 » et de « 439 » à « 576 » : une double ou plusieurs lectures dans les sens inverse (F et R).

Les trois designs sélectionnés et illustrés dans les **figures 11, 12 et 13** ont pu être validés et acceptés. Cependant, leur utilisation s'effectue en fonction de l'échantillon afin de minimiser le nombre de primers tout en gardant une double lecture couvrant la totalité de la séquence étudiée.

2. Chromatogramme : ce résultat permet de visualiser l'alignement et la comparaison de la séquence du contig avec la « rCRS ». Les points de différences entre les deux séquences ne concernent qu'une base (SNP) et sont distingués par des points noirs qui illustrent une transition, une transversion, une délétion, une insertion ou une hétéroplasmie. Chaque différence est vérifiée à partir du chromatogramme où, chaque pic représente une base de la séquence lue. Cette vérification est essentielle afin de confirmer que la variation en question n'est pas due à une erreur de lecture surtout au

voisinage des « Poly C », d'où l'importance d'avoir une double lecture voir plus (plus d'un alignement), tout au long de la séquence.

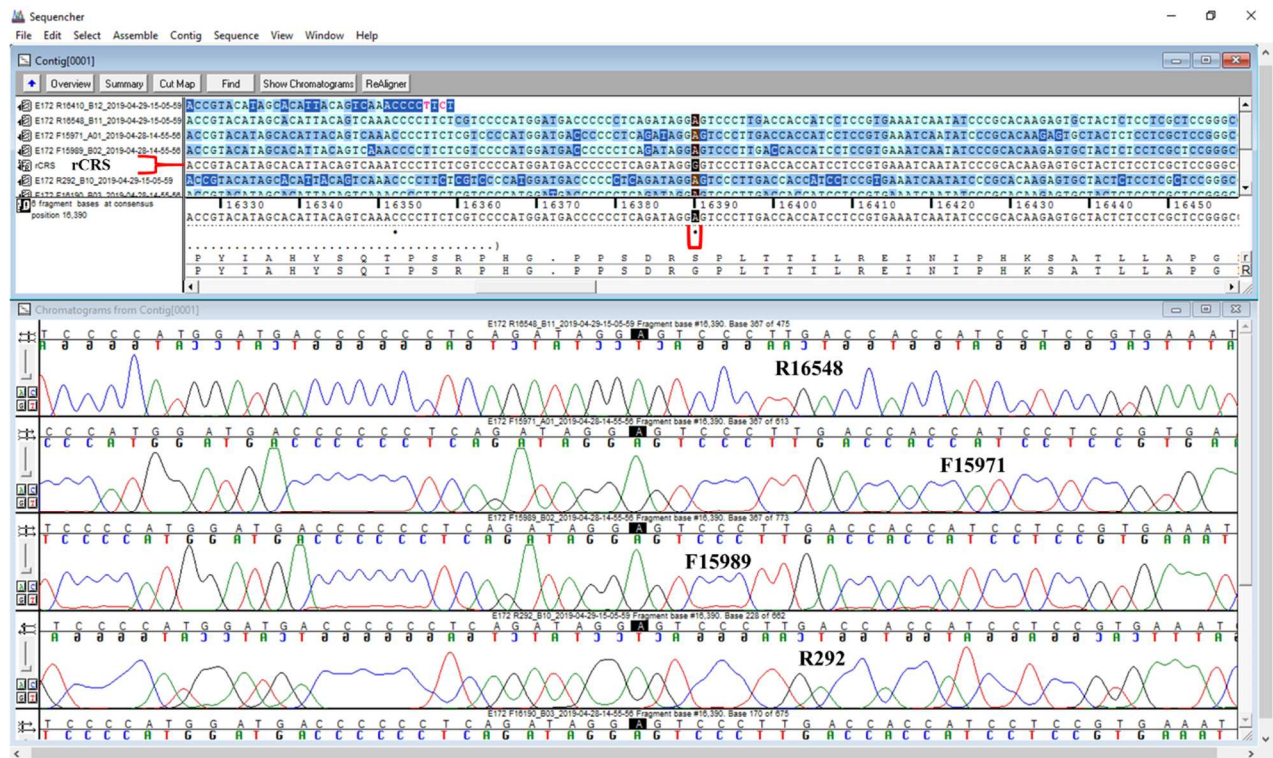


Figure 28. Un alignement d'une séquence d'un échantillon avec la « rCRS ».

Au niveau de la **figure 28** nous pouvons visualiser l'importance d'une double lecture. Comme indiqué, le point noir montre la différence entre la séquence de l'échantillon et celle de référence. Le fait d'avoir deux primers au minimum donnant des séquences ayant la même base, ceci confirme que ce n'est pas une erreur de lecture lors du séquençage mais une variation entre l'échantillon et la « rCRS ». Les quatre électrophorégramme illustre la séquence lue par quatre primers différents et chacune d'elle ayant la même variation par rapport à la séquence référence. Cette variation est une transition de la base G en A.

Les **figures 29** et **30** montrent l'hétéroplasmie de séquence qui est représentée par deux pics différents mais superposables illustrant deux bases différentes. Il est nécessaire d'avoir au minimum une double lecture afin de la confirmer et de la valider par le logiciel « Sequencher® 5.4.2 » qui l'indiquera par une lettre « Y ». Ce polymorphisme aide et rend l'identification plus précise.



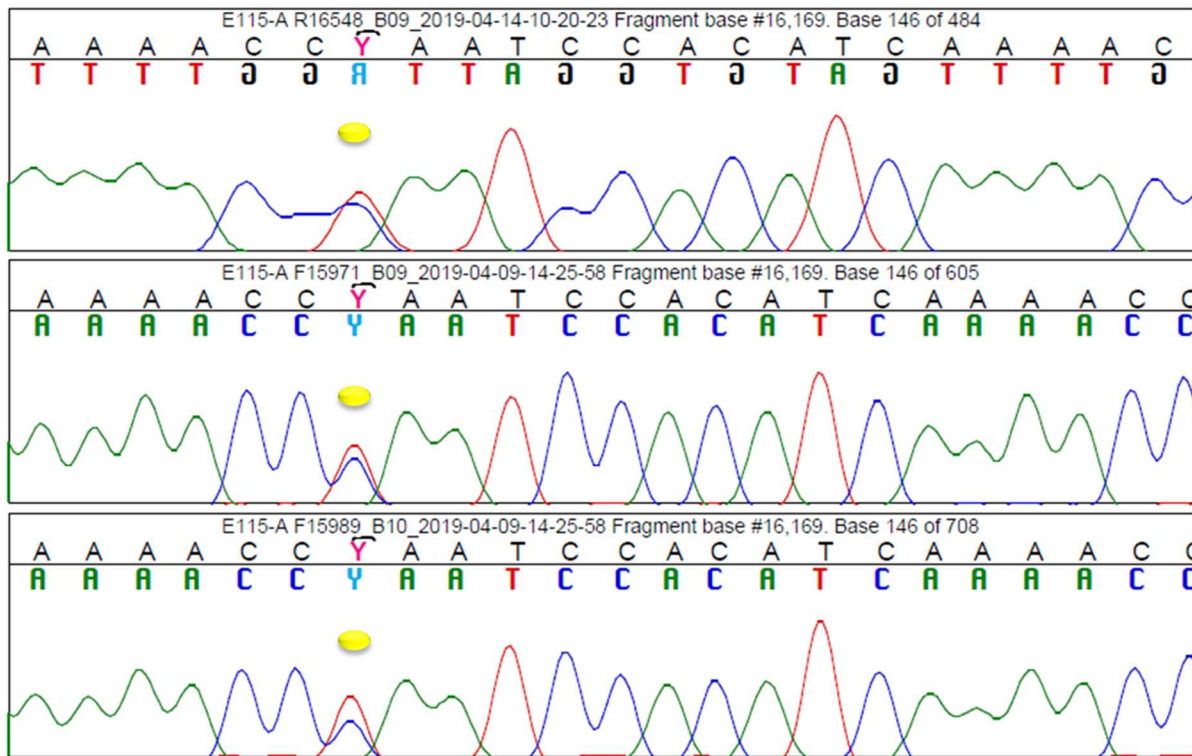


Figure 29. Hétéroplasmie du E115-A.

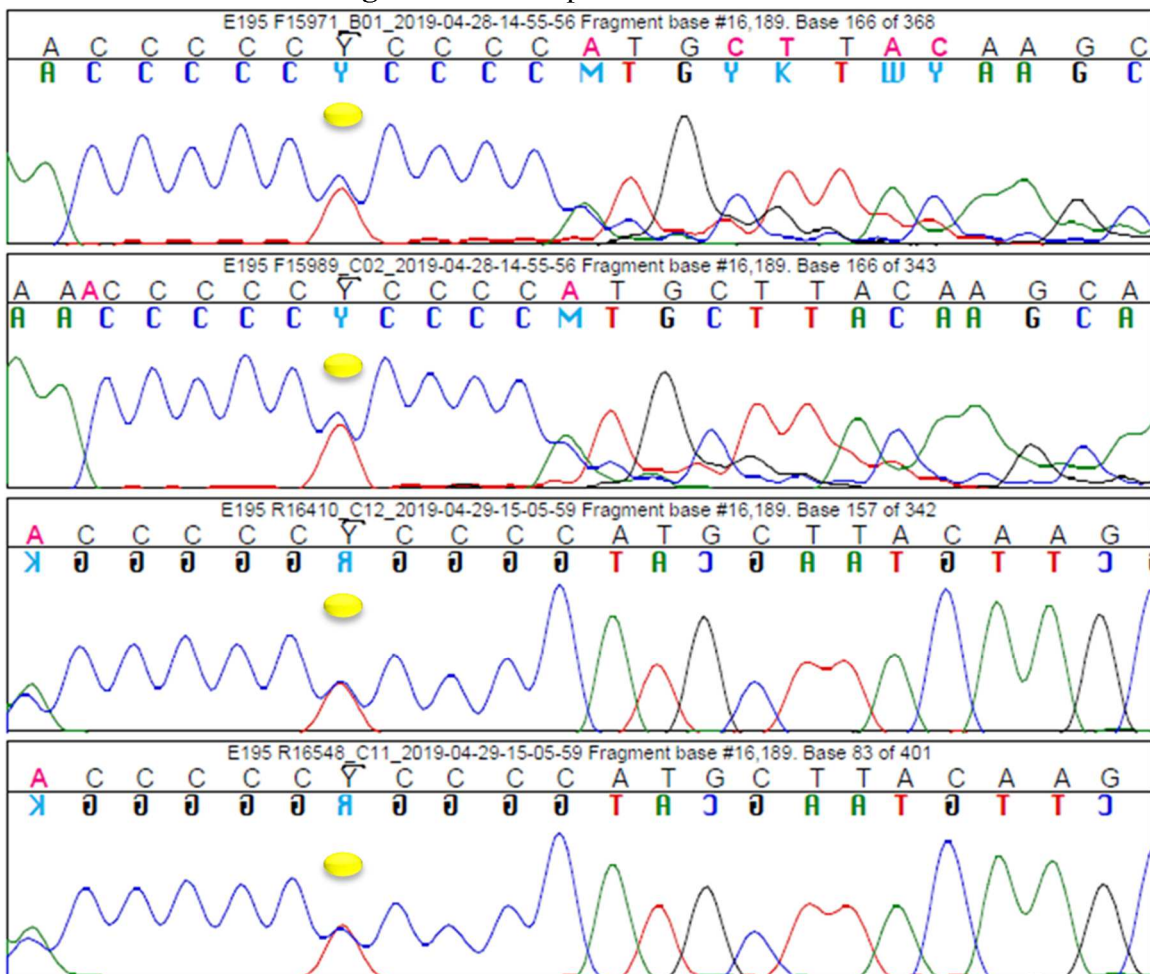


Figure 30. Hétéroplasmie du E195.

3. Rapport du profil ADNmt: à la fin de la visualisation et de la vérification de l'alignement et de la comparaison de la séquence de l'échantillon étudié avec la séquence référence, un tableau englobant toutes les différences visualisées est obtenu.

**Tableau XII.** Résultat du profil du T<sup>+</sup> (9947A).

Référence (rCRS)		T <sup>+</sup> (9947A)
Position	Base	Base variable
16311	T	C
16519	T	C
93	A	G
195	T	C
214	A	G
263	A	G
309.1	:	C (ins)
309.2	:	C (ins)
315.1	:	C (ins)

ins : insertion

« : » : vide

L'ADNmt T<sup>+</sup> (9947A) du « SRM 2392 » a permis de valider notre étude à travers le critère de validation « l'exactitude ». Suite à l'analyse effectuée et après alignement et comparaison à la séquence de référence le profil de cet échantillon est obtenu (**Annexe III**). Le **tableau XII** regroupe toutes les différences entre la « rCRS » et le T<sup>+</sup>, qui se résume en neuf différences. Lors de l'étude du « SRM 2392 », le résultat de l'alignement de l'ADNmt 9947A à la séquence de référence a été similaire au notre (**Annexe III**) ([Levin et al., 2003](#)), y compris pour la dénomination des variations qui s'est faite selon les recommandations.

**Tableau XIII.** Résultat du profil du C<sup>+</sup> (CHR).

Référence (rCRS)		C <sup>+</sup> (CHR)
Position	Base	Base variable
16182.1	:	C (ins)
16183	A	C
16189	T	C
16223	C	T
16278	C	T
16519	T	C
73	A	G
195	T	C
204	T	C
207	G	A
263	A	G
309.1	:	C (ins)
315.1	:	C (ins)

ins : insertion

« : » : vide

L'analyse de l'ADNmt standard C<sup>+</sup> (CHR) a permis de valider notre protocole et les choix des primers. Ceci a été possible en comparant le nombre et le type de différence obtenus dans notre étude avec ceux obtenus dans l'étude du second ADNmt du standard de référence « SRM 2392 » (**Annexe III**). Le profil du C<sup>+</sup> regroupe les 13 différences remarquées en alignant la séquence du cet ADNmt standard avec celle de la « rCRS » (**Annexe III**). Ces variations sont classées dans le **tableau XIII**, ce sont les mêmes retrouvées dans l'étude du « SRM 2392 » ([Levin et al., 2003](#)). Seul point distinct est que la nomination d'une des variations est différente de celle donnée par le Sequencher : 16182.1 est une insertion d'un C alors que la nomination donnée par Levin et al., est noté 16193.1 C selon les recommandations de nomenclatures ([Levin et al., 2003](#)).

### III.3. Résultats de l'étude des témoins négatifs

Le but des témoins négatifs dans notre étude est de vérifier la présence de contamination dans deux étapes clés de l'analyse : l'extraction dont le T<sup>-</sup> est vérifié lors de l'électrophorèse sur gel, et au niveau de l'étape de séquençage où un nouveau T<sup>-</sup> est introduit. Il est vérifié lors de la lecture des données avec le logiciel « Sequencing Analysis Software v6.0 ». Les T<sup>-</sup> ont été utilisés tout au long de notre travail et sur tous nos tests effectués.

En plus de la possibilité de vérifier la contamination lors des manipulations, nous avons effectué une étude sur les T<sup>-</sup> introduit lors du séquençage. Toutes les données concernant l'intensité du signal des quatre bases (G, A, T, C) ont été récupérées et regroupées dans un tableau (**Annexe III**). Suite à cela une moyenne a été calculée pour chaque base pour tous les T<sup>-</sup> afin de déterminer un seuil d'intensité acceptable à ne pas dépasser et qui renseignera aussi sur le taux de contamination.

Le **tableau XIV** regroupe tous les résultats des calculs effectués sur les T<sup>-</sup>, ainsi que la valeur maximale et minimale pour chaque base. Nous avons préféré ajouter le résultat de la médiane qui a été plus représentatif que celui de la moyenne. De ce fait en prenant compte ses deux dernières, il a été possible de fixer le seuil de l'intensité du signal pour chaque base dans un intervalle : [10-25] RFU.

**Tableau XIV.** Moyennes des valeurs de l'intensité du signal (RFU) de chaque base des témoins négatifs.

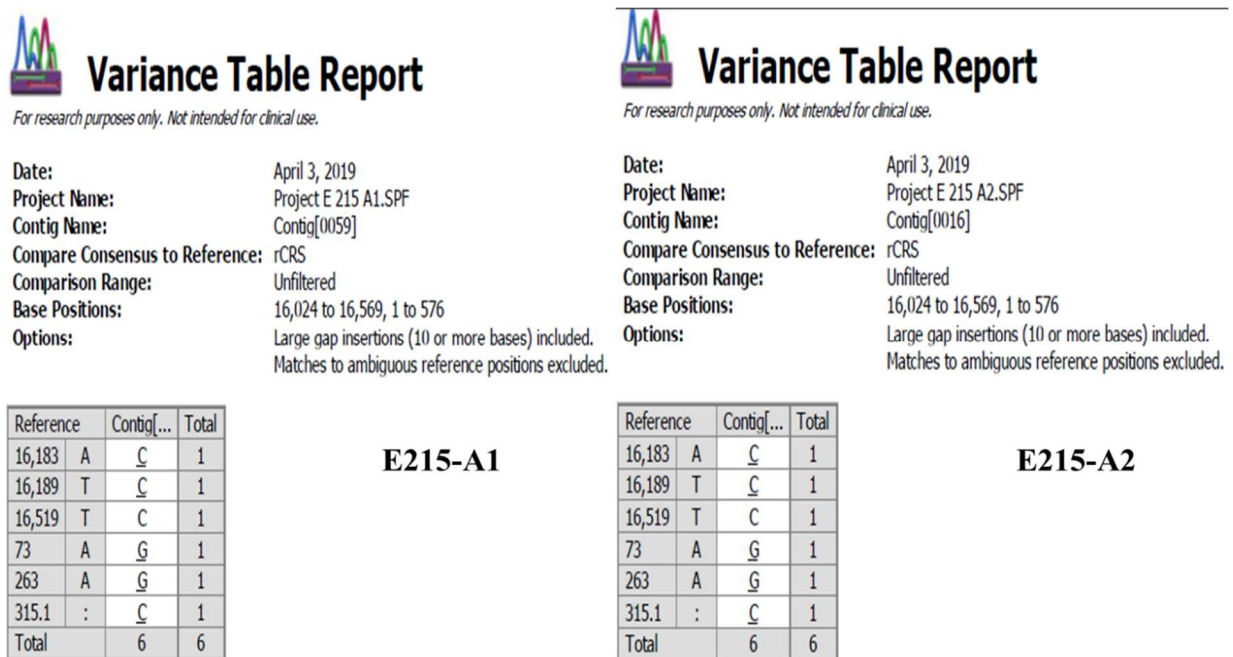
	G	A	T	C
$\bar{x}$	13,216	14,852	31,727	26,614
$\sigma$	6,734	14,545	104,685	71,768
$\bar{x} \pm \sigma$	13,216 $\pm$ 6,734	14,852 $\pm$ 14,545	31,727 $\pm$ 104,685	26,614 $\pm$ 71,768
<b>Médiane</b>	12	12	16	14
<b>Max</b>	62	139	977	656
<b>Min</b>	8	9	13	11

Les trois résultats possibles fournis par le logiciel « Sequencing Analysis Software v6.0 » : le Raw Data, l'électrophorégramme et l'intensité du signal des bases, d'un témoin négatif, ont en commun d'être non interprétable et de faible intensité. Dans certains cas, l'intensité du signal des bases pourrait induire une fausse interprétation en raison de valeurs élevées. Pour cette raison, il est préférable de prendre en considération ces trois résultats.

### III.4. Résultats de la validation

Afin de valider notre protocole trois critères ont été choisis : l’exactitude, la répétabilité et la reproductibilité. Le résultat de chaque critère est le suivant :

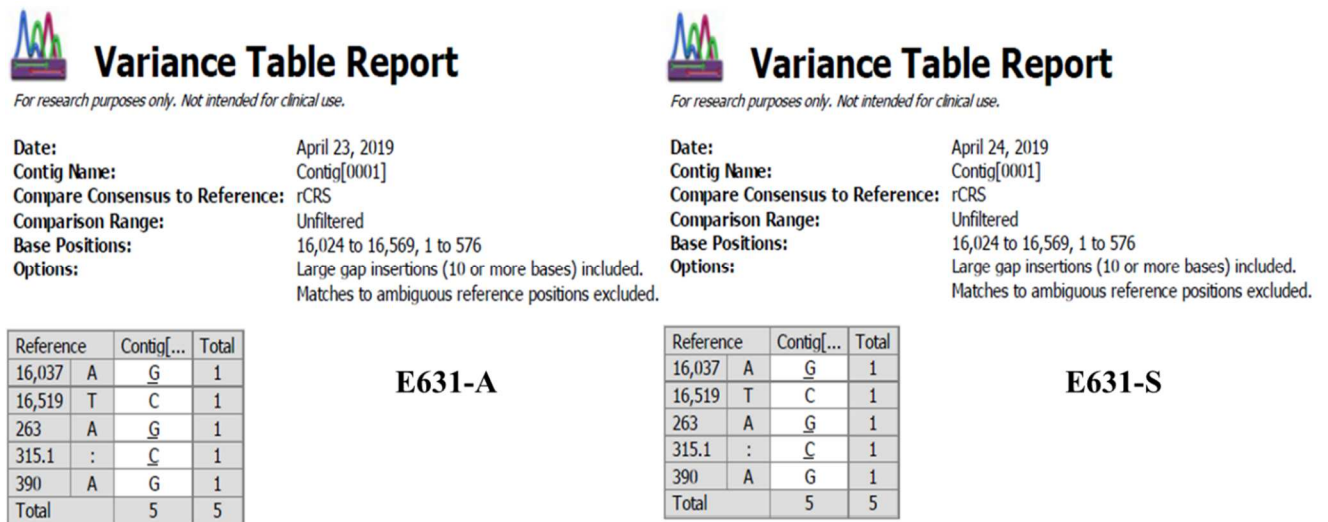
- a. Exactitude : pour la validation de ce critère qui permet de vérifier si les résultats sont similaires ou proches à ceux obtenus lors d’une étude validée, nous avons analysés deux ADNmts standards C<sup>+</sup> et T<sup>+</sup> dont les résultats ont été exposés plus haut.
- b. Répétabilité : ce critère a été appliqué sur la totalité des échantillons analysés. La **figure 31** est un exemple du profil d’un échantillon obtenu en validant ce critère :



**Figure 31.** Profil du E215 validant le critère de répétabilité

Comme la **figure 31** le montre, les deux analyses de cet échantillon ont donné le même profil à savoir six différences au totale.

- c. Reproductibilité : un exemple d'un résultat de validation de ce critère est illustré dans la **figure 32** :



**Figure 32.** Profil du E631 validant le critère de reproductibilité.

Le E631 possèdent 5 variabilités par rapport à la séquence de référence, comme indiqué dans la **figure 32** les deux analyses validant la reproductibilité ont permis d'avoir les mêmes résultats.

Le **tableau XV** est un récapitulatif englobant toutes les variations retrouver dans la « CR » par rapport à la « rCRS » pour tous les échantillons de notre étude. Elles sont citées par ordre de position dans la séquence et noté de la sorte :

- ❖ Substitution : la position suivie de 'base de la rCRS>base de l'échantillon'.
- ❖ Insertion : la position avant l'insertion suivie d'un point et nombre '1' de la 1<sup>ère</sup> ou l'unique insertion et puis de la lettre de la base ajouter.
- ❖ Délétion : la base de la rCRS suivie de la position (s) et enfin Del.

Tableau XV. Récapitulatif des profils des échantillons étudiés.

T+	C+	M+	E9	E10	E18	E115	E172	E195	E215	E231	E234	E630	E631
													16037 A>G
			16069 C>T	16069 C>T									
										16114 C>A			
			16126 T>C	16126 T>C		16126T>C							
												16129G>A	
							16145 G>A						
					16153 G>A								
						16169 C>Y							
							16176 C>G						
	16182.1C												
												16182 A>G	
	16183 A>C								16183 A>C			16183 A>C	
											16186 C>T		
	16189T>C							16189 T>Y	16189 T>C	16189 T>C			
			16193 C>T										
												16198 T>C	
	16223 C>T						16223 C>T						
				16241 A>G									

T+	C+	M+	E9	E10	E18	E115	E172	E195	E215	E231	E234	E630	E631
												16249 T>C	
										16256 C>T			
										16270 C>T			
	16278 C>T		16278 C>T										
						16294 C>T				16294 C>T			
					16298T>C			16298 T>C					
						16304 T>C					16304 T>C		
16311 T>C												16311 T>C	
							16352 T>C						
							16390 G>A						
16519 T>C	16519 T>C	16519 T>C				16519 T>C	16519T>C	16519 T>C	16519 T>C			16519 T>C	16519 T>C
										16526 G>A			
					72 T>C			72 T>C					
	73 A>G		73 A>G	73 A>G		73 A>G	73 A>G		73 A>G	73A>G		73 A>G	
93 A>G													
							106-111Del						
			150 C>T	150 C>T									
		152 T>C	152 T>C				152 T>C						
			185 G>A										
195T>C	195T>C			195 T>C	195 T>C		195 T>C					195 T>C	
	204T>C												



T+	C+	M+	E9	E10	E18	E115	E172	E195	E215	E231	E234	E630	E631
	207G>A												
214A>G													
263A>G	263A>G	263A>G	263A>G	263A>G	263A>G	263A>G	263A>G	263A>G	263A>G	263A>G	263A>G	263A>G	263A>G
			295C>T	295C>T									
309.1C	309.1C		309.1C	309.1C	309.1C					309.1C	309.1C		
309.2C													
315.1C	315.1C	315.1C	315.1C	315.1C	315.1C	315.1C	315.1C	315.1C	315.1C	315.1C	315.1C	315.1C	315.1C
													390A>G
				456C>T							456C>T		
		477T>C											
						482T>C							
			489T>C	489T>C								489T>C	
								513 G>A					
			523 DelA				523 DelA						
							524 DelC						
				524.1 A									
			525 DelC										
				525.1 C									

16069 C>T : substitution d'une base C par T à la position 16069

Y : hétéroplasmie

315.1 C : insertion de la base C à la position 315

523 DelA : délétion d'une base A à la position 523

### III.5. Discussion Générale

L'ADNmt représente un intérêt majeur dans l'identification génétique humaine du fait de plusieurs caractéristiques (Cf. Chap. I, p. 8) dont le taux de mutations qui est supérieur à l'ADN nucléaire, particulièrement au niveau de la D-loop ou région de contrôle. En raison de l'absence du système de réparation et de recombinaison, la « CR » possède, au niveau de ses régions hypervariables, un nombre élevé de différences qui varient d'un individu à un autre constituant ainsi, un mitotype discriminant et fiable, ce qui fait de cette région un choix prometteur qui donnerais des résultats pointilleux (Lembring, 2013).

Au cours de notre étude pour l'élaboration du protocole d'analyse de l'ADNmt, plusieurs essais dont le choix des primers selon trois stratégies distinctes et l'ajout d'une étape de pré-séquençage, ont été réalisés sur différents échantillons nous permettant ainsi d'améliorer le séquençage de la CR et de s'assurer de son bon déroulement malgré la nature variante des régions polycytosines pour chaque échantillon.

En effet, les profils obtenus lors de cette analyse montrent que les régions hypervariables de la D-Loop sont riches en variations SNPs et que leur utilisation en criminalistique apparaît évidente. Pour la comparaison des mitotypes, le séquençage de la HVIII en plus de la HVI et HVII est nécessaire. Lutz et *al.*, 2000 ont démontré que dans certains cas, les séquences analysées à partir de HVI et HVII peuvent être identiques entre deux individus, alors qu'en incluant l'analyse de HVIII, le profil devient plus discriminant en révélant de nouvelles variations au niveau de cette région, d'où l'intérêt d'analyser et de séquencer les trois régions hypervariables de la « CR ».

La visualisation et la comparaison des mitotypes obtenus nous a permis de constater plusieurs similitudes entre les variations trouvées dans notre travail et celles démontrées dans des études précédentes telles que celles de Brandstätter et *al.* en 2004 et Chaitanya et *al.*, en 2016. Ces similitudes concernent particulièrement les deux variations : 263G et 315.1C qui ont été retrouvées dans tous les mitotypes de nos échantillons, appuyant ainsi notre protocole.

De plus, nous avons constaté dans les profils obtenus que la plupart des substitutions sont de types transition (purine par purine ou pyrimidine par pyrimidine), ce qui concorde avec ce qui a été rapporté par les travaux de Tully et *al.* en 2001.

L'analyse de l'ADNmt est sensible et minutieuse exigeant différentes recommandations que ça soit pour la manipulation (Cf. Chap II, p. 16) ou pour l'établissement d'un profil. Les variations trouvées dans ce dernier doivent être dénommées suivant les recommandations mise en place par EDNAP et SWGDAM afin d'éviter toute ambiguïté dans l'interprétation et la comparaison des profils avec des banques de données.

Concernant la nomenclature des insertions, il serait préférable de noter : la position qui vient juste avant l'insertion, suivie d'un point et du nombre '1' pour la 1<sup>ère</sup> base ajoutée, et continuer de façon successive pour chaque base insérée par ex : 309.1C, 309.2C. Lorsque l'insertion se situe au niveau des « Poly C » de HVI et HVII la nomenclature de l'insertion est généralement effectuée à partir de la dernière position du bloc de l'homopolymère, par ex pour HVI : 16193.1C. Alors que pour HVII elle est généralement, considérée en deux blocs c'est-à-dire

303-309 et 311-315 de ce fait la nomenclature sera : 309.1C et 315.1 C respectivement. Pour ce qui est des délétions, elles doivent être rapportées à la position supprimée par ex : Si la position entre 524 et 526 a été retirée elle sera notée : 525Del. Enfin, la substitution est souvent notée selon trois possibilités soit : la base de référence puis la position suivie de la base variable « A263G », soit seulement la position et la base différente « 263G » ou directement la nomenclature « 263 A>G ». L'hétéroplasmie de séquence est généralement notée soit par une lettre Y dans le cas où l'hétéroplasmie est bien distinguée ou par la lettre N en cas d'ambiguïté de lecture (Tully et *al.*, 2001 ; Wilson et *al.*, 2002 ; Parson et *al.*, 2014 ; SWGDAM, 2014 ; SWGDAM, 2019).

Le profil obtenu suite à l'analyse de l'ADNmt, renseigne sur l'identité génétique des individus et sur leur filiation maternelle. De ce fait, en criminalistique ce type de profil est plus spécifique à la lignée qu'à l'individu lui-même, ce qui pourrait créer une confusion lors de l'identification génétique. Dans le cas de comparaison de deux mitotypes, la présence de la même hétéroplasmie, renforce la spécificité de l'identité génétique, du fait que ce type de polymorphisme est un évènement rare (Tully et *al.*, 2001).

# **Conclusion et perspectives**

## **Conclusion et Perspective**

Notre étude a été réalisée dans le but d'établir et de valider un protocole d'analyse des régions hypervariables de l'ADNmt référence pour l'identification d'individus en forensique, essentiellement pour pallier aux soucis de l'état de l'échantillon d'ADN à analyser.

En premier lieu et à travers ce travail, il a été possible de mettre au point les recommandations et les précautions lors de l'analyse de l'ADNmt, ainsi que de sélectionner et de classer les différentes étapes et procédures de cette analyse. En second lieu, au cours des différents essais, il a été nécessaire de rajouter des étapes (lavage) ou d'en retirer d'autres (dilution), contribuant ainsi à l'amélioration du protocole.

Un des points importants de cette étude a été l'analyse de plusieurs échantillons issus de prélèvements de salive afin de mettre en place les différents designs possibles. Ceci a eu pour objectif d'avoir une analyse et un séquençage des régions hypervariables de la « CR » qui soient fait de la meilleure façon donnant ainsi un résultat précis, en fonction de l'échantillon présentant ou non un Homopolymère « polyC ».

Autre point crucial est que ce protocole a mis en exergue différents équipements et réactifs performants dont l'utilisation a été un avantage pour l'accomplissement de l'analyse dans les bonnes conditions et l'obtention de profils sans aucun problème d'interprétation.

De plus, les profils obtenus ont permis de mettre en évidence le pouvoir discriminant des polymorphismes présents dans l'ADNmt et l'importance du choix des régions hypervariables de la D-Loop (CR) pour l'identification génétique. Aussi, les mitotypes des échantillons analysés ont permis d'appuyer et de valider notre protocole du fait de la similitude des types de variations obtenus avec différentes études déjà réalisées sur l'ADNmt à travers le monde.

Par ailleurs, ce protocole donne la possibilité à d'autres travaux sur l'ADNmt. Ainsi, il serait possible d'analyser différents échantillons d'individus de différentes régions algériennes, de les comparer et de les introduire à des bases de données et pouvoir établir les haplogroupes de la population et de mettre en évidence ses origines, et par la suite contribuer à la conception d'une base de données de la population algériennes.

Entre autres, l'utilisation de l'ADNmt ne se limite pas uniquement à la science forensique et sa manipulation peut être utile et importante dans d'autres domaines. En effet, en prenant en compte l'importance de l'ADN ancien pour les sciences anthropologiques et la sensibilité de l'ADNnc il n'est pas toujours facile de l'utiliser dans ce domaine d'étude. De ce fait, il serait prometteur de proposer une optimisation du protocole d'analyse de l'ADNmt étant plus résistant et mieux conservable et de contribuer ainsi à son utilisation dans les études anthropologiques, permettant l'avancée des travaux concernant l'histoire, et l'ascendance de la population algérienne essentiellement du point de vue biologique.

Enfin il serait aussi intéressant d'améliorer le protocole et de l'affiner pour d'autres types d'échantillons biologique (sang, cheveux, ossement, etc.) et de nature peu avantageuse (dégradé, faible quantité, etc.) afin de pouvoir l'utiliser dans de différents cas réels à savoir établir des liens de parenté de lignées maternelles, identification d'individus et essentiellement de victimes

de tragédies extrême au point d'affecter l'ADNnuc et de permettre ainsi l'avancement du domaine criminalistique en Algérie.

# **Références bibliographiques**

## Références bibliographiques

### A

**Anderson A.**, Bankier A.T., Barrell B.G., de Bruijn M.H.L., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreir P.H., Smith A.J.H., Staden R. et Young I.G., (1981), Sequence and organisation of the human mitochondrial genome, *Nature*, Vol : 290, pp 457-465

**Andrews R.M.**, Kubacka I., Chinnery P.F., Lightowlers R.N., Turnbull D.M. et Howel N., (1999), Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nature Genetics*, Vol : 23, pp 147

Applied Biosystems BigDye® XTerminator™ Purification Kit, Protocol, (2007)

Applied Biosystems Sequencing Analysis Software Quick Reference Card, (2009)

### B

**Bandelt H.J.**, Brandstätter A.K., Richards M.B., Yao Y.G. et Logan I., (2014), The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies, *Journal of Human Genetics*, Vol : 59, pp 66-77

**Bell J.R.**, (2008), A Simple Way to Treat PCR Products Prior to Sequencing Using ExoSAP-IT®, USB Corporation, *Biotechniques*, Vol : 44, No : 6, pp 834

**Bell J.**, (1989), The polymerase chain reaction, *Immunology Today*, Vol : 10, No : 10, pp 351-355

**Berger C.**, Hatzler-Grubwieser P., Hohoff C., Parson W., (2011), Evaluating sequence-derived mtDNA length heteroplasmy by amplicon size analysis, *Forensic Science International: Genetics*, Vol : 5, pp 142–145

**Blanchin M.D.**, (2010), Validation des méthodes d'analyse

**Boujard D.** (2015), *Biologie Cellulaire et Moléculaire*, 2e édition, Paris, DUNOD : pp 85

**Brandstätter A.**, Peterson C.T., Irwin J.A., Mpoke S., Koech D.K., Parson W., Parsons T.J., (2004), Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database, *Int J Legal Med*, Vol : 118, pp 294-306

**Budowle B.**, Allard M.W., Wilson M.R. et Chakraborty R.(2003), *Forensics and mitochondrial DNA: Applications, Debates, and Foundations Annual Review of Genomics and Human Genetics*, Vol : 4, pp 119-141

**Budowle B.**, Schutzer S.E., Morse S.A., Martinez K.F., Chakraborty R., Marrone B.L., Messenger S.L., Murch R.S., Jackson P.J., Williamson P., Harmon R., et Velsko S.P., (2008), Criteria for Validation of Methods in Microbial Forensics, *Applied And Environmental Microbiology*, Vol. 74, No. 18 pp : 5599-5607



**Butler J.M.**, (2005), Forensic DNA Typing, Biology, Technology, and Genetics of STR Markers 2<sup>nd</sup> edition, USA Elsevier, 647 p.

**Butler J.M.**, (2011), Advanced Topics in Forensic DNA Typing : Methodology, USA, Elsevier, 680 p.

**Butler J.M.**, (2010), Fundamentals of Forensic DNA Typing, USA, Elsevier, 485 p

**Butler J.M.**, (2014) Validation Overview, National Institute of Standards and Technology

**Butler J.M.**, Coble. M.D, Vallone. P.M, (2007), STRs vs. SNPs: thoughts on the future of forensic DNA testing, Forensic Sci Med Pathol, Vol : 3, pp 200-205

## C

**Carracedo A.** (2005), Forensic DNA Typing Protocol, New Jersey, Humana Press, 279 p

**Chaitanya L.**, Van Oven M., Brauer S., Zimmermann B., Huber G., Xavier C., Parson W., de Knijff P., Kayser M., (2016), High-quality mtDNA control region sequences from 680 individuals sampled across the Netherlands to establish a national forensic mtDNA reference database, Forensic Science International: Genetics, 36 p

**Coudray C.**, Torroni A., Achilli A., Pala M., Olivieri A., Larrouy G., Dugoujon J.M., (2009), Les lignées mitochondriales et l'histoire génétique des populations berbérophones du nord de l'Afrique, Antropo, Vol : 18, pp 63-72

## D

**Dumache R.**, Ciocan V., Muresan C.O. et Enache A., Molecular Genetics and its Applications in Forensic Sciences *in*, Shetty. B.S et Padubidri. J.R, (2016), Forensic Analysis : From Death to Justice, Romania, pp 88-96

## E

**Edlund H.**, (2010), Sensitive Identification Tools in Forensics DNA Analysis, Digital comprehensive summaries of Uppsala dissertations from the Faculty of Medicine 660, Uppsala, 49 p

**Elrod S.** et STANSFIELD W., (2003), Génétique, 4e édition, USA, Ediscience, pp 429-431

ExoSAP-IT USB corporation, (2000), Molecular Systems and F. Hoffmann-La Roche.

## F

**Falkenberg M.**, (2018), Mitochondrial DNA replication in mammalian cells: overview of the pathway, Essays in Biochemistry, Vol : 62, pp 287-269

**Friedman J.R.** et Nunnari J., (2014), Mitochondrial form and function, Nature, Vol : 505, pp 335-343.

## G

**Garibyan L.** et Avashia N., (2013), Research Techniques Made Simple: Polymerase Chain Reaction (PCR), *J Invest Dermatol*, Vol : 133, No : 3, pp 1-8

**Goodwin W.**, Linacre A. et Hadi S., (2007), *An Introduction to Forensic Genetics*, UK, WILEY, pp 151.

## H

**Hlinka V.**, Muharam I. et Ientile V.K., Method validation in forensics and the archaeological sciences *in*, Haslam M., Robertson G., Crowther A., Nugent S., Kirkwood L., (2009), *Archaeological Science Under a Microscope: Studies in Residue and Ancient DNA Analysis in Honour of Thomas H. Loy*, Australie, ANU E Press, 278 p

**Holland M.M.** et Parsons T.J., (1999), Mitochondrial DNA Sequence Analysis-Validation and Use for Forensic Casework, *Forensic Science Review*, Vol : 11, No : 1, pp 22-50

**Huber N.**, Parson W., Dür A., (2018) Next generation database search algorithm for forensic mitogenome analyses, *Forensic Science International: Genetics*, Vol : 37, pp 204-214

## I

**Ishmael F.T.** et Stellato C., (2008), Principles and applications of polymerase chain reaction: basic science for the practicing physician, *Annals Of Allergy, Asthma & Immunology*, Vol : 101, pp 437-443

**Ivanov P.L.**, Wadhams M.J, Roby R.K., Holland M.M., Weedn V.W., et Parsons T.J., (1996), Mitochondrial DNA sequence heteroplasmy in the Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II, *Nature Genetics*, Vol : 12, pp 417-420

## J

**Jeffreys A.J.** (2005), Genetic Finger Printing, *Nature Medicine*, Vol :11, pp 1035-1039.

## K

**Kivisild T.**, (2015), Maternal ancestry and population history from whole mitochondrial genomes, *Investigative Genetics*, Vol : 6, No : 3, pp 1-10

**Kobilinsky L.**, Liotti F.T. et Oeser-Sweat J., (2005), *DNA : Forensic and Legal Applications*, New Jersey, WILEY, 347 p

**Kong, Y.**, Trabucco, S. E. et Zhang, H., (2014), Oxidative stress, mitochondrial dysfunction and the mitochondria theory of aging, *Interdisciplinary topics in gerontology*, Vol : 39, pp 86-107

## L

**Ladoukakis E.D.** et Zouros E., (2017), Evolution and inheritance of animal mitochondrial DNA: rules and exceptions, *Journal of Biological Research-Thessaloniki*, Vol : 24, No : 2, pp 1-7

**Lembring M.**, (2013), application of mitochondrial DNA analysis in contemporary and historical samples. Digital comprehensive summaries of Uppsala dissertations from the Faculty of Medicine 950, Uppsala, 62 p

**Levin B.C.**, Hancock D.K., Holland K.A., Cheng H., et Richie K.L., (2003), Human Mitochondrial DNA- Amplification and Sequencing – Standard Reference Materials-SRM 2392 and SRM 2392-I, NIST Special Publication 260-155, 33 p

**Lutz S.**, Wittig H., Weisser H.J., Heizmann J., Junge A., Dimo-Simonin N., Parson W., Edelmann J., Anslinger K., Jung S. et Augustin C., (2000), Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Science International Vol : 113 pp 97-101

## N

**Nicholls T.J.** et Minczuk M., (2014), In D-loop: 40 years of mitochondrial 7S DNA, Experimental Gerontology, Elsevier, Vol : 56, pp 175-181.

## P

**Pakendorf. B** et Stoneking. M, (2005), Mitochondrial DNA and human evolution, Annual Review of Genomics and Human Genetics, Vol : 6, pp 165-183

**Panneerchelvam S.** et Norazmi M.N., (2003), FORENSIC DNA PROFILING AND DATABASE, Malaysian Journal of Medical Sciences, Vol : 10, No : 2, pp 20-26

**Pham-Hoia E.**, Pussiaua A., Curcia J.Y., (2017), ADN, une perpétuelle avancée, Revue Francophone Des Laboratoires, Criminalistique, No : 489, pp 67-76

**Parson W.**, Brandstätter A., Alonso A., Brandt N., Brinkmann B., Carracedo A., Corach D., Froment O., Furach I., Grzybowski T., Hedberg K., Keyser-Tracqui C., Kupiec T., Lutz-Bonengel S., Mevag B., Ploski R., Schmitter H., Schneider P., Syndercombe-Court D., Sørensen E., Thew H., Tully G., Scheithauer R., (2004), The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives, International Congress Series, Vol : 139, pp 215-226

**Parson W.**, Brandstätter A., Pircher M., Steinlechner M., Scheithauer R., (2004), EMPOP—the EDNAP mtDNA population database concept for a new generation, high-quality mtDNA database, International Congress Series, Vol : 1261, pp 106-108

**Parsons L.**, et Bright J.A., (2012), A manual and automated method for the forensic analysis of DNA from buccal samples on Whatman Indicating FTA Elute Cards, Australian Journal of Forensic Sciences, Vol : 44, No : 4, pp 393-402

**Parson W.**, Gusmao L., Hares D.R., Irwin J.A., Mayr W.R., Morling N., Pokorak E., Prinz M., Salas A., Schneider P.M., Parsons T.J., (2014), DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing, Forensic Science International: Genetics, Vol : 13, pp 134-142

**Parson W.**, Dür A., (2007), EMPOP- a Forensic mtDNA database, Forensic Science International: Genetics, Vol : 1, pp 88-92

**Prainsack B.**, Aronson J.D., Forensic Genetic Databases: Ethical and Social Dimensions *in*, Wright D.J., (2015), International Encyclopedia of the Social & Behavioral Sciences, 2<sup>ème</sup> édition, USA, Elsevier, pp 339-345

**Prainsack B.**, Cole S.A., Genetics and Forensics *in*, Wright D.J., (2015), International Encyclopedia of the Social & Behavioral Sciences, 2<sup>ème</sup> édition, USA, Elsevier, pp 955-961

**Prieto L.**, Zimmermann B., Goios A., Rodriguez-Monge A., Paneto G.G., Alves C., Alonso A., Fridman C., Cardoso S., Lima G., Anjos M.J., Whittle M.R., Montesino M., Cicarelli R.M.B., Rocha A.M., Albarran C., de Pancorbo M.M., Pinheiro M.F., Carvalho M., Sumita D.R., Parson W., (2011), The GHEP-EMPOP collaboration on mtDNA population data-A new resource for forensic casework, Forensic Science International: Genetics, Vol : 5, pp 146-151

PROTOCOL SUMMARY BigDye® XTerminator™ Purification Kit, (2006)

## R

**Reid R.**, (2018), Applications of the mitochondrion in Forensic DNA typing *in*, Dash. H.R, Shrivastava. P, Mohapatra. B.K and Das. S, (2018), DNA Fingerprinting: Advancements and Future Endeavors, Singapore, Springer, pp 241-255

**Runa D.** et Walsh S.J., (2006), The Continuing Evolution of Forensic DNA Profiling-From STRS to SNPS, Australian Journal of Forensic Sciences, Vol : 38No : 2, pp 59-74

## S

**Samehsalari S.** et Reddy K.R., (2018), Application Control Region of Human Mitochondrial DNA in Forensic Anthropology, International Journal of Modern Anthropology, Vol 2(11), pp 233-248.

**Sato M.** et Sato K., (2013), Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA, Biochimica et Biophysica Acta, Vol : 1833, pp 1979–1984

Scientific Working Group on DNA Analysis Methods, (2014), Mitochondrial DNA Nomenclature Examples Document

Scientific Working Group on DNA Analysis Methods, (2016), Validation Guidelines for DNA Analysis Methods, pp 1-15

Scientific Working Group on DNA Analysis Methods, (2019), Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories

Sequencher® 5 Series, User Manual, Gene Codes Corporation, (2015)

**Shen M.** et Vieira D.N., (2016), Forensic science: defending justice, Forensic Sciences Research, Vol : 1, No : 1, pp 1-2

**Shewale J.G** et Liu R.H., (2016), ForensicDNA AnalysisCurrent Practices and Emerging Technologies, USA, Taylor & Francis Group, pp 381

Standard for Internal Validation of Forensic DNA Analysis Methods, (2018), 1<sup>ère</sup> édition, ASB Standard 038, pp 1-4

**Stewart J.B.** et Chinnery P.F., (2015), The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease, *Nature Genetics*, Vol : 16, pp 530-542

## T

**Taanman J.W.**, (1999), *The Mitochondrial Genome : Structure, Transcription, Translation and Replication*, *Biochimica et Biophysica, Acta*, Elsevier, pp 103-123.

**Tully G.**, Bär W., Brinkmann B., Carracedo A., Gill P., Morling N., Parson W., Schneider P., (2001), Consideration by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles, *Forensic Science International*, Vol : 124, pp 83-91

**Turner P.C.**, Mc Lennan A.G., Bates A.D., et White M.R.H., (2006), *L'essentiel en Biologie Moléculaire*, 2e édition, UK, BERTI Edition, pp 187-190

## U

**Umetsua K.** et Yuasab I., (2005), Recent progress in mitochondrial DNA analysis, *Legal Medicine*, Vol : 7, pp 259-262

User Bulletin, Sequencing Analysis Software v5.3, (2007)

## V

**Van Oven M.** et Kayser M., (2008), Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation, *Human Mutation*, Vol : 30, pp 386-394

**Valko M.**, Leibfritz D., Moncol J., Cronin MT.D., Mazur M. et Telser J., (2007), Free radicals and antioxidants in normal physiological functions and human disease », *The International Journal of Biochemistry & Cell Biology*, Vol : 39, No : 1, pp 44-84.

## W

**Wallace DC.**, Brown M.D., Lott MT., (1999), Mitochondrial DNA variation in human evolution and disease, *Gene*, Vol : 238, pp 211–230

**Wallace D.C.** et Chalkia D., (2013), Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease, *Cold Spring Harb Perspect Biology*, Vol : 5, pp 1-47

**Wilson M.R.**, Allardb M.W., Monsona K., Millerc K.W.P., Budowle B., (2002), Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region, *Forensic Science International*, Vol : 129, pp 35-42

## Y

**Yasukawa T.** et Kang D., (2018), An overview of mammalian mitochondrial DNA replication mechanisms, *The Journal of Biochemistry*, Vol : 164, No : 3, pp 183-193

# **Annexes**

## Annexes

### Annexe I

Séquence de la « CR ».

16024 TTCTTTC ATGGGGAAGC AGATTGGGT ACCACCCAAG TATTGACTCA CCCATCAACA  
16081 ACCGCTATGT ATTCGTACA TTA CTGCCAG CCACCATGAA TATTGTACGG TACCATAAAT  
16141 ACTTGACCAC CTGTAGTACA TAAAAACCCA ATCCACATCA AAACCCCCTC CCCATGCTTA  
16201 CAAGCAAGTA CAGCAATCAA CCCTCAACTA TCACACATCA ACTGCAACTC CAAAGCCACC  
16261 CCTCACCCAC TAGGATACCA ACA AACCTAC CCACCCTTAA CAGTACATAG TACATAAAGC  
16321 CATTTACCGT ACATAGCACA TTACAGTCAA ATCCCTTCTC GTCCCATGG ATGACCCCCC  
16381 TCAGATAGGG GTCCCTTGAC CACCATCCTC CGTGAAATCA ATATCCCGCA CAAGAGTCT  
16441 ACTCTCCTCG CTCCGGGCC ATAACACTTG GGGGTAGCTA AAGTGAAGT TATCCGACAT  
16501 CTGGTTCTTA CTT CAGGTC ATAAAGCCTA AATAGCCCAC ACGTTCCCCT TAAATAAGAC  
16561 ATCACGATG  
1 GATCACAGGT CTATCACCT ATTAACCACT CACGGGAGCT CTCCATGCAT TTGGTATTTT  
61 CGTCTGGGGG GTATGCACGC GATAGCATTG CGAGACGCTG GAGCCGGAGC ACCCTATGTC  
121 GCAGTATCTG TCTTTGATTG CTGCCTCATC CTATTATTA TCGCACCTAC GTTCAATATT  
181 ACAGGCGAAC ATACTTACTA AAGTGTGTTA ATTAATTAAT GCTTGTAGGA CATAATAATA  
241 ACAATTGAAT GTCTGCACAG CCACTTTCCA CACAGACATC ATAACAAAAA ATTTCCACCA  
301 AACCCCCCTC CCCCCGCTTC TGGCCACAGC ACTTAAACAC ATCTCTGCCA AACCCCCAAA  
361 ACAAGAACC CTAACACCAG CTAACCCAGA TTTCAAATTT TATCTTTTGG CGGTATGCAC  
421 TTTAACAGT CACCCCCCAA CTAACACATT ATTTCCCCT CCCACTCCCA TACTACAAAT  
481 CTCATCAATA CAACCCCCGC CCATCCTACC CAGCACACAC ACACCGCTGC TAACCCATA  
541 CCCC GAACCA ACCAAACCC AAAGACACCC CCCACA

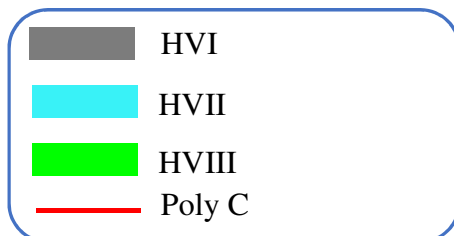


Figure 33. Séquence de la région de contrôle (D-loop)

## Annexe II

**Tableau XVI. Matériels non Biologique**

Etapes du protocole	Appareillage	Equipements de manipulation	Réactifs	Equipements de protection
<b>Echantillonnage</b>		Carte FTA, Ecouvillon buccale .		
<b>Extraction</b>	Centrifugeuse, Agitateur, UV crosslinker, BioCap.	stylo micro-emporte-pièce (Harris UniCore Punch), Tapis à découpe, Portoirs, Micro-tubes 0.2ml, Tubes 5ml, Micropipette (100µl, 200µl) et Tips.	Tampon TE « Tris, EDTA ».	
<b>Amplification</b>	Vortex, Centrifugeuse, Thermocycleur, Hotte à flux laminaire, UV crosslinker.	Micropipette (10µl, 100µl, 200µl) Tips, Portoirs, Micro-tubes 0.2ml et 1.5ml.	AmpliTaq Gold 360 Master Mix, BSA, Amorces F et R, Eau ultra-pure.	
<b>Quantification sur Gel</b>	Balance électronique, Microonde, Système de migration, Générateur de courant.	Bécher 1000ml, Erlenmeyer 250ml et Eprouvette graduée 50ml, Tips, Micropipette (100µl), Moule et peigne.	TAE « Tris, Acétate, EDTA », Agarose, SYBR, Blue Juice, Low DNA Mass Ladder, Eau ultra-pure.	Gants, Masque d'hygiène, Charlotte.
<b>Purification « ExoSap »</b>	Vortex, Thermocycleur, Hotte à flux laminaire, UV crosslinker.	Micropipette (10µl), Tips, Micro-tubes 0.2ml, portoirs.	ExoSap.	
<b>Séquençage</b>	Vortex, Thermocycleur, Hotte à flux laminaire, UV crosslinker.	Micropipette (10µl, 100µl, 200µl) Tips, Micro-tubes 0.2ml et 1.5ml, portoirs.	Ready Reaction Mix, Sequencing Buffer, Amorces, Eau ultra-pure.	
<b>Purification « Big-dye X-Terminator »</b>	Vortex, Agitateur, Centrifugeuse, Hotte à flux laminaire, UV crosslinker, Automate sous-vide.	Micropipette (1000µl, 200µl, 100µl et 10µl), Tips, Plaque 96 puits, Support de Plaque, Portoirs, Film transparent, Micro-tubes 1.5ml ou 2ml.	Solution X-Terminator, Solution SAM, Eau ultra-pure.	



<b>Electrophorèse Capillaire</b>	Séquenceur 3500xl, UV crosslinker.	Septa, Plaque pour séquenceur.	
<b>Lecture des Données</b>	Logiciels : Sequencing Analysis, Sequencher.		

**Tableau XVII.** Les amorces utilisées en séquençage

<b>Amorces</b>	<b>Séquence</b>	<b>Taille</b>	<b>Brin</b>
<b>F15971</b>	5'-TTAACTCCACCATTAGCACC-3'	20 nucléotides	Sens
<b>F15989</b>	5'-CCCAAAGCTAAGATTCTAAT-3'	20 nucléotides	Sens
<b>F16190</b>	5'-CCCCATGCTTACAAGCAAGT-3'	20 nucléotides	Sens
<b>F15</b>	5'-CACCTATTAACCACTCACG-3'	20 nucléotides	Sens
<b>F403</b>	5'-TCTTTTGGCGGTATGCACTTT-3'	21 nucléotides	Sens
<b>F456</b>	5'-CCCCTCCCACTCCCATACT-3'	19 nucléotides	Sens
<b>R614</b>	5'-TTTCAGTGTATTGCTTTGAGGA-3'	22 nucléotides	Antisens
<b>R568</b>	5'-GTGTCTTTGGGGTTTGGTTG-3'	20 nucléotides	Antisens
<b>R292</b>	5'-ATTTTTTGTTATGATGTCT	19 nucléotides	Antisens
<b>R16548</b>	5'-GGGAACGTGTGGGCTATTTA-3'	20 nucléotides	Antisens
<b>R16410</b>	5'-GAGGATGGTGGTCAAGGGAC-3'	20 nucléotides	Antisens

## Annexe III

CRS		C+	T+	Region
# <sup>a</sup>	Base <sup>b</sup> 1981/1999	Template CHR <sup>d</sup>	Template 9947A <sup>d</sup>	
73	A	G	-	HV2
93	A	-	G	HV2
150	C	-	-	HV2
152	T	-	-	HV2
185	G	-	-	HV2
195	T	C	C	HV2
204	T	C	-	HV2
207	G	A	-	HV2
214	A	-	G	HV2
228	G	-	-	HV2
263*R	A	G	G	HV2
295	C	-	-	HV2
303-309	-	C (ins)	CC (ins)	HV2
311-315*R	-	C (ins)	C (ins)	HV2
462	C	-	-	HV2
482	T	-	-	HV2
489	T	-	-	HV2
16069	C	-	-	HV1
16126	T	-	-	HV1
16183	A	C	-	HV1
16184-93	-	C (ins)	-	HV1
16189	T	C	-	HV1
16193	C	-	-	HV1
16223	C	T	-	HV1
16278	C	T	-	HV1
16292	C	-	-	HV1
16311	T	-	C	HV1
16357	T	-	-	HV1
16362	T	-	-	HV1
16519	T	C	C	HV1

**Figure 34.** Différences entre la « rCRS » et l'ADNmt des deux échantillons du « SRM 2392 » au niveau des régions hypervariables (non-codantes) (Levin et al., 2003)



# Variance Table Report

*For research purposes only. Not intended for clinical use.*

**Date:** March 24, 2019  
**Contig Name:** Contig[0001]  
**Compare Consensus to Reference:** rCRS  
**Comparison Range:** Unfiltered  
**Base Positions:** 16,024 to 16,569, 1 to 576  
**Options:** Large gap insertions (10 or more bases) included.  
Matches to ambiguous reference positions excluded.

Reference		Contig[...]	Total
16,311	T	<u>C</u>	1
16,519	T	C	1
93	A	<u>G</u>	1
195	T	<u>C</u>	1
214	A	<u>G</u>	1
263	A	<u>G</u>	1
309.1	:	<u>C</u>	1
309.2	:	<u>C</u>	1
315.1	:	<u>C</u>	1
Total		9	9

**Figure 35.** Profil du T<sup>+</sup> (9947A).



# Variance Table Report

*For research purposes only. Not intended for clinical use.*

**Date:** April 3, 2019  
**Project Name:** Project C+1 UNFINISHED.SPF  
**Contig Name:** Contig[0019]  
**Compare Consensus to Reference:** rCRS  
**Comparison Range:** Unfiltered  
**Base Positions:** 16,024 to 16,569, 1 to 576  
**Options:** Large gap insertions (10 or more bases) included.  
Matches to ambiguous reference positions excluded.

Reference		Contig[...]	Total
16,182.1	:	<u>C</u>	1
16,183	A	<u>C</u>	1
16,189	T	<u>C</u>	1
16,223	C	<u>I</u>	1
16,278	C	<u>I</u>	1
16,519	T	C	1
73	A	<u>G</u>	1
195	T	<u>C</u>	1
204	T	<u>C</u>	1
207	G	<u>A</u>	1
263	A	<u>G</u>	1
309.1	:	<u>C</u>	1
315.1	:	<u>C</u>	1
Total		13	13

**Figure 36.** Profil du C<sup>+</sup> (CHR).